DAD: DISTRIBUTIVE ANALYSIS / ANALYSE DISTRIBUTIVE

USER'S MANUAL

Université Laval
MIMAP PROJECT
DAD4.2
The software DAD4.2 was designed by Jean-Yves Duclos and
Araar Abdelkrim and programmed with JDK1.4
by Araar Abdelkrim and Carl Fortin
June 2002

Jean-Yves Duclos : jyves@ecn.ulaval.ca
Abdelkrim Araar : aabd@ecn.ulaval.ca
Carl Fortin : cfortin@gel.ulaval.ca

Université Laval

Introduction

DAD was designed to facilitate the analysis and the comparisons of social welfare, inequality, poverty and equity across distributions of living standards. Its features include the estimation of a large number of indices and curves that are useful for distributive comparisons as well as the provision of asymptotic standard errors to enable statistical inference. The features also include basic descriptive statistics and provide simple non-parametric estimations of density functions and regressions.
The main facilities of DAD are the:

1. Estimation of indices of:
   - Poverty (Watts, CHU, FGT, S-Gini. Sen): normalised and un-normalised (or absolute and relative poverty indices), with absolute and relative poverty lines
   - Social Welfare (Atkinson, S-Gini, Atkinson-Gini)
   - Inequality (S-Gini, Atkison, Entropy, Atkinson-Gini and others)
   - Redistribution, progressivity, vertical equity, reranking and horizontal inequity.

2. Decomposition of:
   - Poverty across population subgroups
   - Inequality across population subgroups or by "factor components" (e.g., by type of consumption expenditures or source of income)
   - Progressivity and equity across different taxes and/or tranfers and subsidies
   - Poverty changes across growth and redistribution effects.

3. Checks for the robustness of distributive comparisons.
4. Estimation of stochastic dominance curves of the primal and dual types, for poverty, social welfare, inequality and equity dominance.
5. Robustness of decompositions into population subgroups and factor components.
6. Estimation of popular "dual" curves: ordinary and generalised Lorenz curves, Cumulative Poverty Gap curves, quantile curves, normalised quantile curves, poverty gap curves, ordinary and generalised concentration curves.
7. Estimation of popular "primal" curves: cumulative distribution functions, poverty deficit curves, poverty depth curves, etc…
8. Estimation of differences in curves and indices.
9. Estimation of "critical" poverty lines for absolute and relative poverty comparisons.
10. Estimation of crossing points for dual curves.
11. Provision of asymptotic standard deviations on all estimates of indices, points on curves, critical poverty lines, crossing points, etc…, allowing for dependence or independence in the samples being compared. These standard deviations are currently computed under the assumption of identically and independently distributed sample observations, but the computations take into account the randomness of the sampling weights when such weights are provided by the user.
12. Allowance for sampling errors in the poverty lines specified to compute absolute and relative poverty indices.

DAD's environment is user-friendly and uses menus to select the variables and options needed for all applications. The software can load simultaneously two data bases, can carry out applications with only one data base or two, and can allow for dependence or independence of data bases and vectors of living standards in computing standard errors on differences in indices and curves.
The databases can be built with the software or can be loaded from a hard disk or a floppy or CD-ROM driver. The databases can be edited, new observations can be added, and new vectors of data can be generated using arithmetical or logical operators.

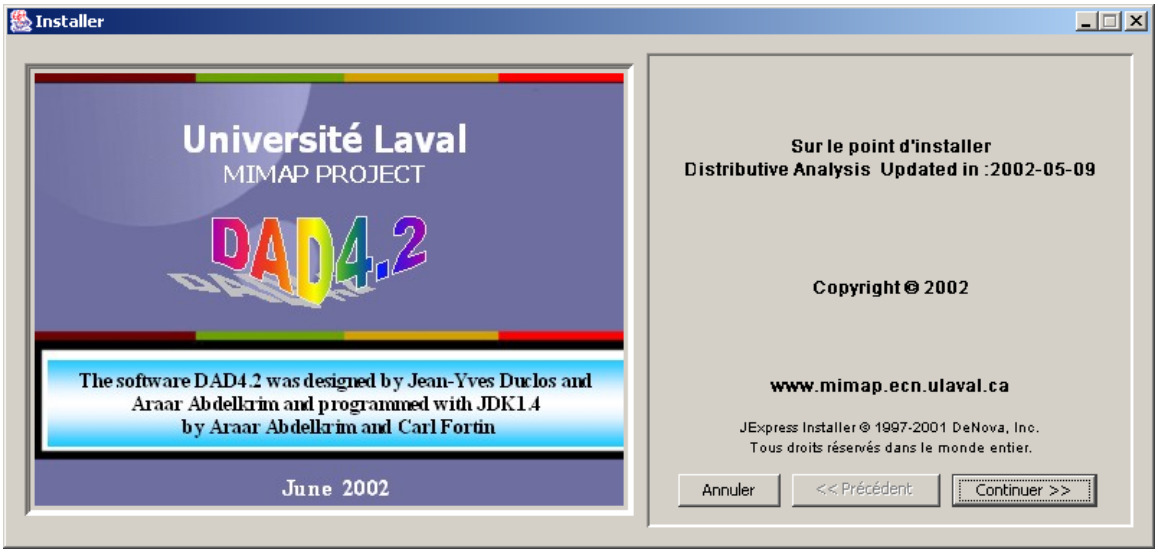
Features of version 4.2 of DAD

- A new specific format to saving and load data in DAD
- Provision of a new output window that adds significantly to the amount of information provided and results in a higher quality of output display.
- A new window to edit the results that can then be saved in HTML format.
- Estimation of new indices and curves and addition of new options for the estimation of indices and curves.
- This version, compiled with JDK 1.4, can run on Windows 95, 98, 2000 and Widows XP.
- More effective data handling, resulting in better memory use and increased capacity to deal with large data bases.
- Optimized algorithms for processing data, yielding a much-increased speed of execution for several computations.

Installation and required equipment

DAD is conceived to run on operating systems Windows 95-98 NT, Windows2000 and Windows XP. A PC of 100MHz or more is also required. The steps for installation of this software are as follows:

1. Insert the CD-ROM that contains the DAD installation file and click on the icon "**jinstall**". The following window appears:


Installer
Université Laval
MIMAP PROJECT
DAD4.2
The software DAD4.2 was designed by Jean-Yves Duclos and Araar Abdelkrim and programmed with JDK1.4 by Araar Abdelkrim and Carl Fortin
June 2002
Sur le point d'installer
Distributive Analysis Updated in :2002-05-09
Copyright © 2002
www.mimap.ecn.ulaval.ca
JExpress Installer © 1997-2001 DeNova, Inc.
Tous droits réservés dans le monde entier.
Annuler
<< Précédent
Continuer >>

Click on the button "**continue**" and specify the installation directory.

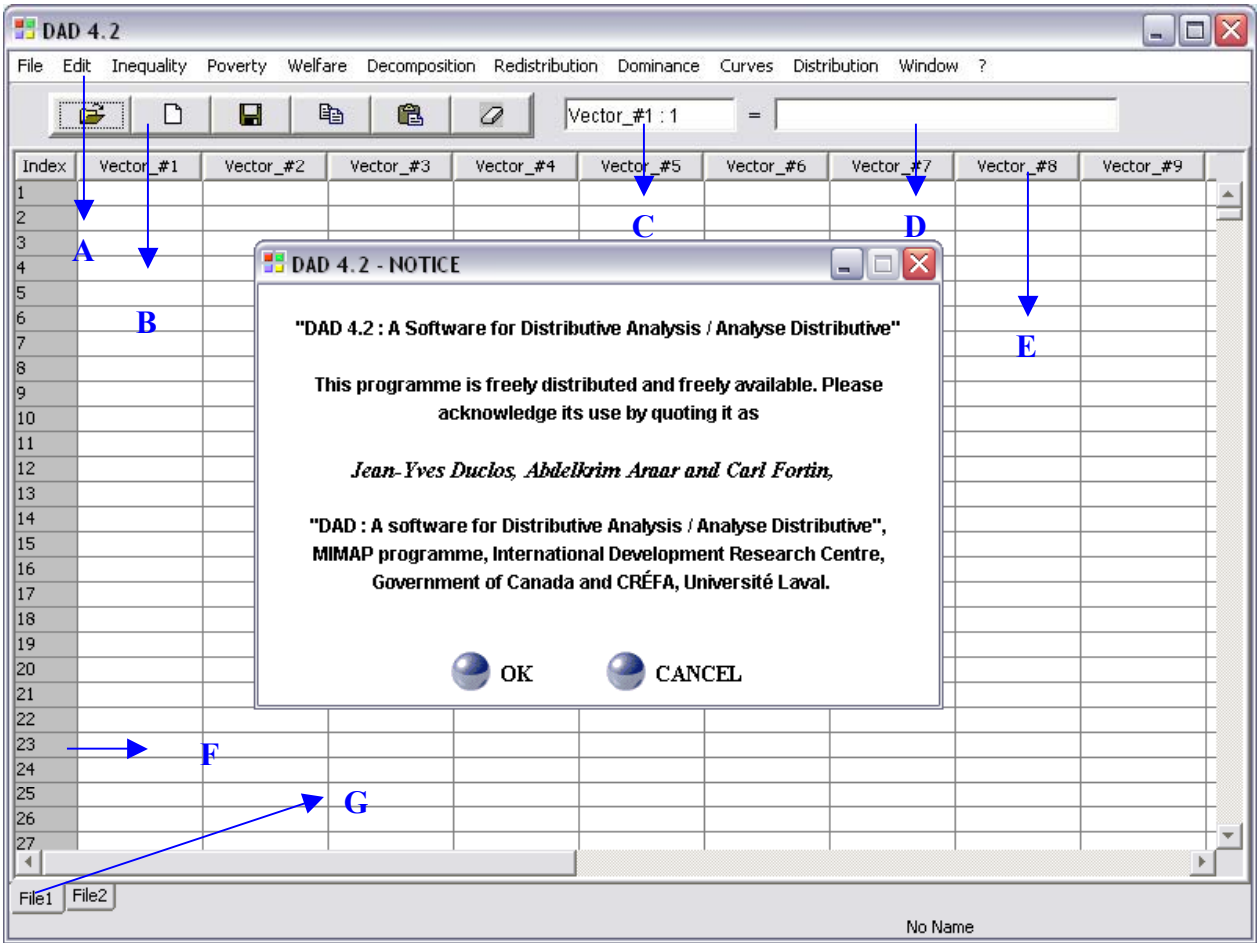
At the end of the procedure of installation, you can run this software like any other program by clicking on the button "**Start**" and selecting the item "Program ⇒ Distributive Analysis ⇒ DAD4.2"

Databases in DAD4.2

A database used in DAD is a set of vectors of data. Each vector represents a specific variable. By default, the length of each vector determines the number of observations for that variable. Each database contains a set of vectors whose number of observations must be the same.

Constructing a database with DAD

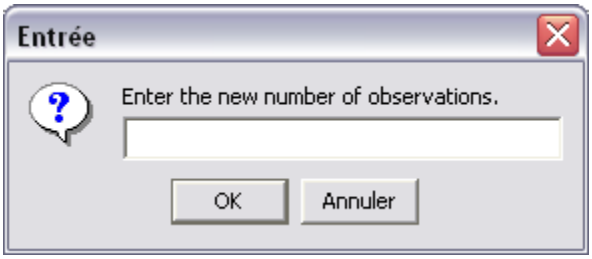
After opening DAD, the following window appears:


DAD 4.2
File Edit Inequality Poverty Welfare Decomposition Redistribution Dominance Curves Distribution Window ?
Vector_#1 : 1 =
Index Vector_#1 Vector_#2 Vector_#3 Vector_#4 Vector_#5 Vector_#6 Vector_#7 Vector_#8 Vector_#9
A
B
C
D
E
F
G
DAD 4.2 - NOTICE
"DAD 4.2 : A Software for Distributive Analysis / Analyse Distributive"
This programme is freely distributed and freely available. Please acknowledge its use by quoting it as
Jean-Yves Duclos, Abdelkrim Araar and Carl Fortin,
"DAD : A software for Distributive Analysis / Analyse Distributive", MIMAP programme, International Development Research Centre, Government of Canada and CRÉFA, Université Laval.
OK
CANCEL
File1
File2
No Name

A- Main menu;
B- Toolbar;
C- The selected cell;
D- Value of the selected cell;
E- Name of column;
F- Index of observation;
G- The selected file.

To construct a new database with DAD, follow these steps:

1. In the main menu, click on the command **"File"** and select the option **"New File".** A window asks the user to indicate the desired number of observations for the new file:



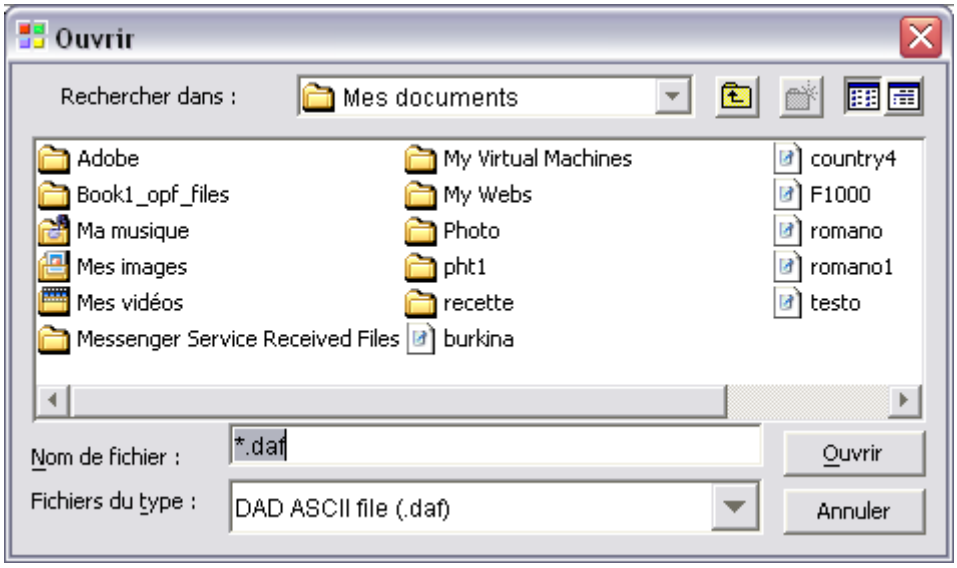
2. Enter the number of observations of the new file and click on the button OK. To begin editing the new vectors, follow these steps:
3. Click on the cell (vector #1, index=1). The contour of this cell changes to yellow.
4. Write the new value of the cell. As a general rule with DAD, the decimal part should be separated by a dot (.).
5. Press **"Enter.**
6. Write the value of the next cell and repeat the procedure until all of values of vector #1 are registered.
7. To edit another vector, select the first cell of this vector and repeat steps 3 up to 6.

If you want to modify the value of any one cell, follow these steps:

1. Select the cell subject to be modified by clicking on it.
2. Write the new value of the cell.
3. Press **"Enter".**

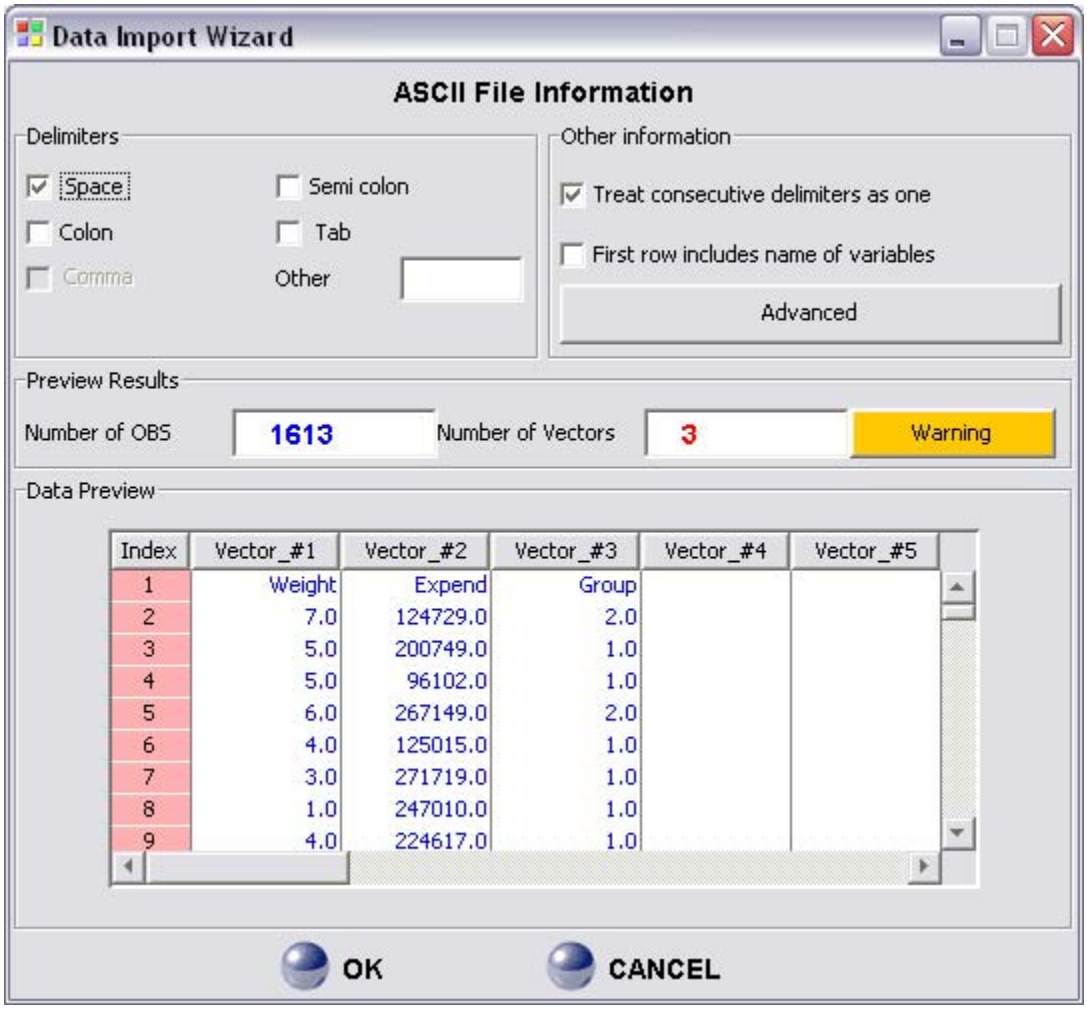
Loading an ASCII data base

To load an ASCII data file, click on the command **"File"**, select the command **"Open".** The following window appears, asking for some information concerning the data file.



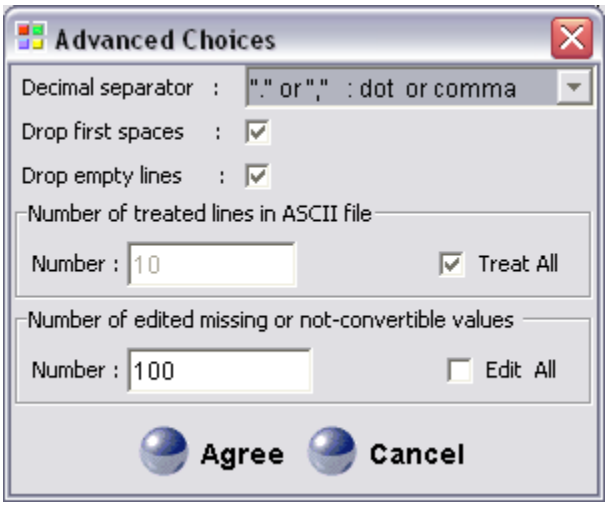
Remark: if your ASCII file's extension is not .txt, .dat, or .prn, choose ".*" in the option "Type of File", then indicate the file name.*

After choosing the desired ASCII file and clicking on OK, the following window appears.



These windows contain many options that facilitate the loading of an ASCII file. By default the delimiter (the character that separates variables) is a space, but you can specify other delimiters. You can also specify the delimiter with the option "Other". In the Panel "Other Information", you can indicate the following information:

1- By default, the option "Treat consecutive delimiters as one" is selected. Choosing this option makes it such that several succeeding delimiters are treated as one.
2- By default, the option "First row includes names of variables" is not selected. In this example, the ASCII file's first row includes the names of variables; we thus select the option.
3- Clicking on the button "**Advanced**" makes the following windows appear:



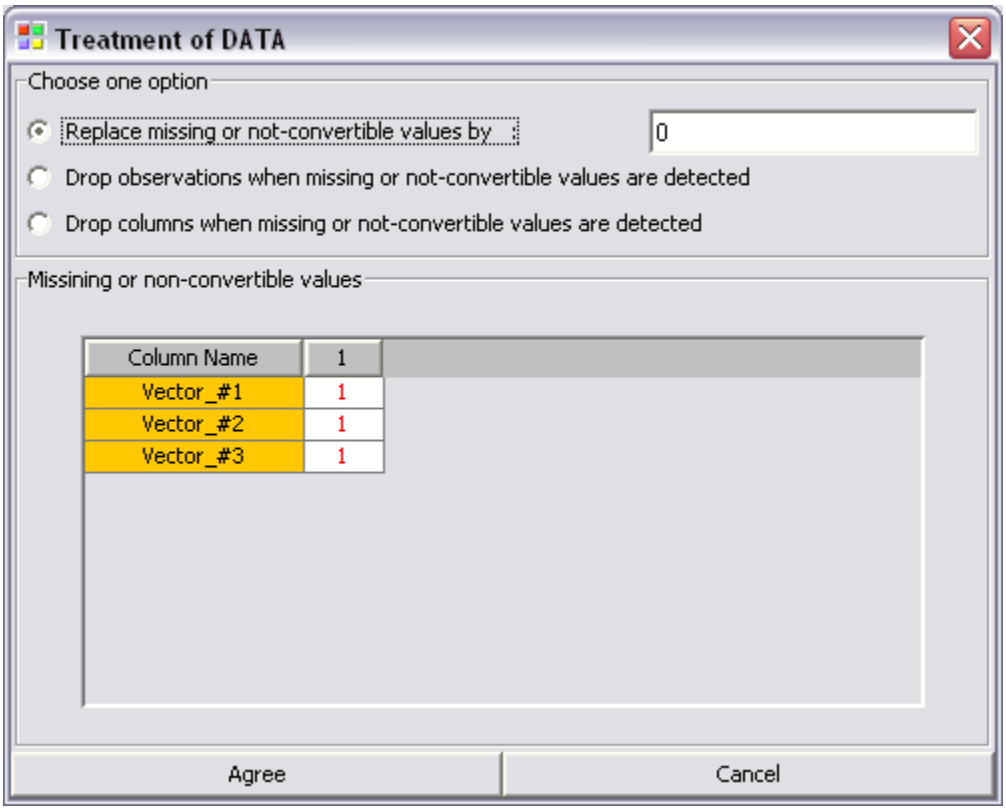
We do not by default need to specify what the separator of decimals is, but if we indicate that it is a dot, then we may specify that the separator between the variables can be a comma.

Remark: If the delimiter of columns is a comma, the delimiter of decimals cannot also be a comma.

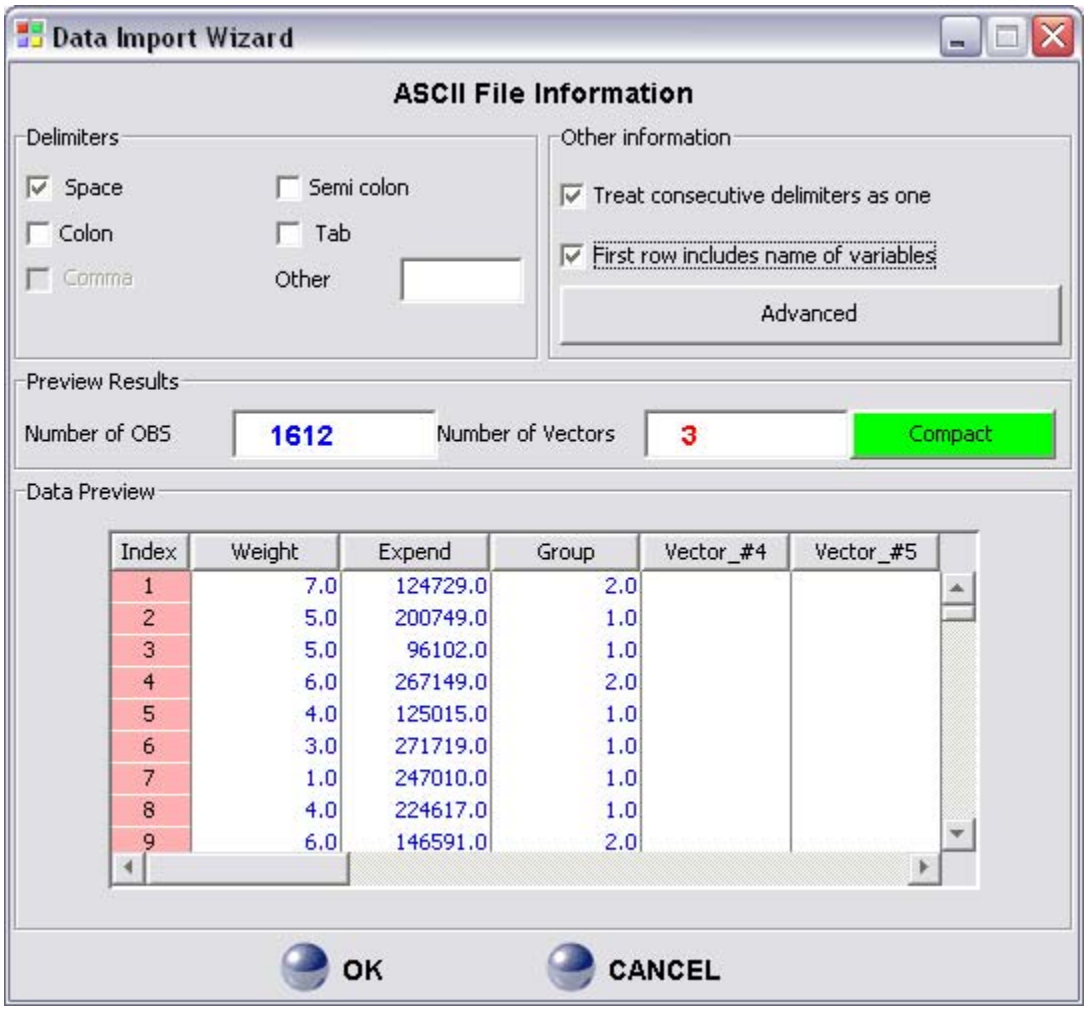
By selecting the option "Drop first spaces", we do not take into account spaces which precede the values of the first column. We can also indicate the number of lines in the ASCII file to be treated, as well as the number of missing or not-convertible values to be edited.

The panel "Preview results" shows the number of observations and the number of columns in the ASCII file. The panel "Data Preview" displays instantaneously the data as their reading changes according to selected options. This a useful tool for reliable loading of ASCII data files.

Note in the panel "Preview Results" the message Button "Warning". If we click on the button, the following window appears:



In the panel "Choose one option" there are three options to treat missing or not convertible values. In our example, we would just indicate that the first row includes the names of variables. Hence, we click on the button "cancel" and we indicate this.



After selecting the option "First row includes names of variables", the button "Compact" replaces the button "Warning". This button indicates that all values in the three columns are acceptable to DAD. At this stage, you can click on the button "ENTER" to finalize the loading of the data.

*Remark: after loading the ASCII file we can save this file with the DAD ASCII format *.daf.*

Loading a second ASCII database

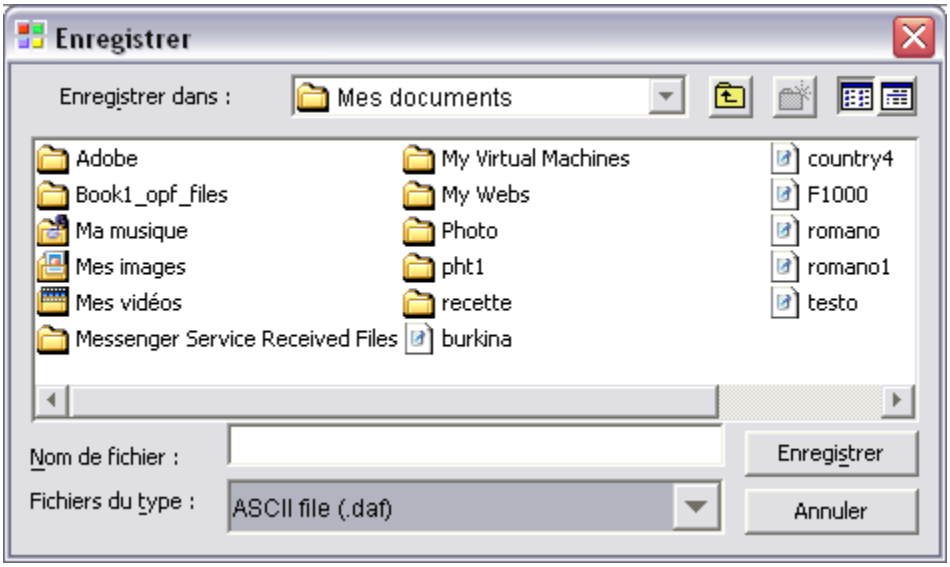
As already mentioned, for many applications in DAD we can use simultaneously two databases. To activate a second database, the user should load another file. To activate a second database, follow these steps:

1. Activate the second file by clicking on the button "File2".
2. The procedures to follow after this are identical to those presented for loading the first ASCII file.

Remark: The "active" file in the software DAD is the selected file.

Loading a DAD ASCII format file

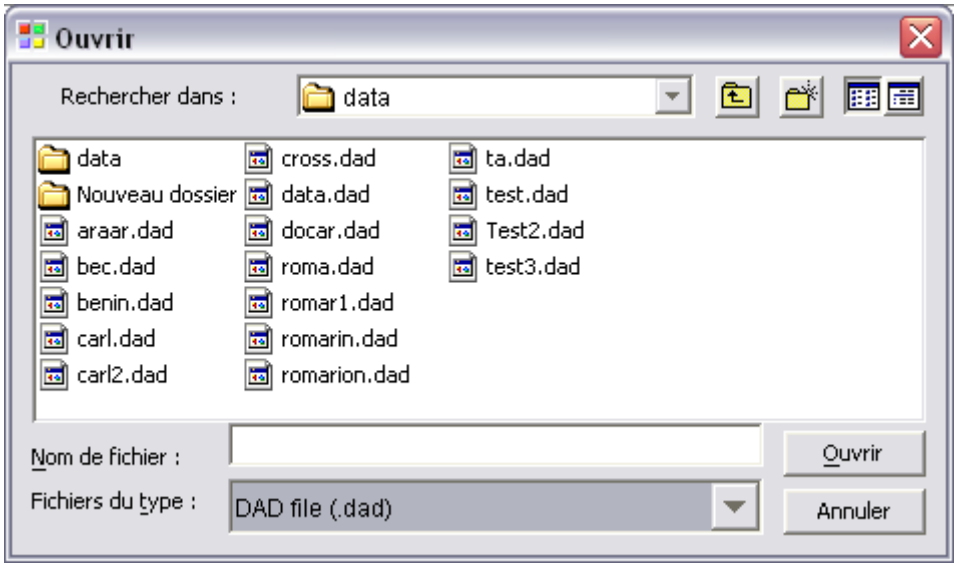
With DAD, you can also save and load files in DAD's specific format and with the extension "*.daf". To open a ".daf" file, click on the command **"File"** and select the command **"Open".** The following window appears, asking for some information concerning the data file.



After this, select the file type "DAD file "(*.daf)", select the file, and click on the Button "Open".

Loading a DAD file

With DAD, you can also save and load files in DAD's specific format and with the extension "*.dad". To open a ".dad" file, click on the command **"File"** and select the command **"Open".** The following window appears, asking for some information concerning the data file.

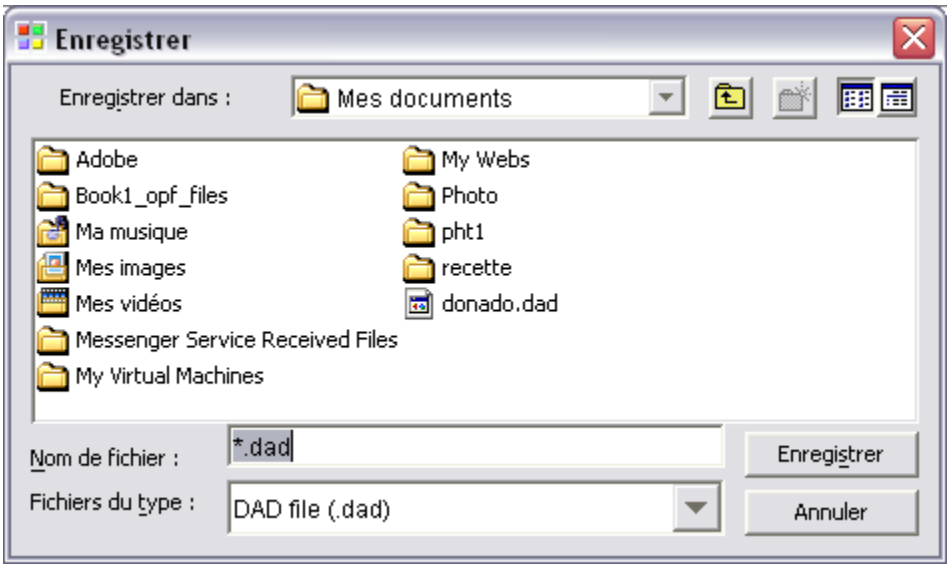


After this, select the file type “DAD file “(*.dad)”, select the file, and click on the Button “Open”.

Remark: DAD files contain two sheets, such as “File1” and “File2”, with every sheet containing one database. It is possible that one of the two sheets be empty.

Saving a file

You can save an active file in DAD’s file format (*.daf or *.dad). The procedure is simple. Begin with the command **"File"** and select the item **"Save".** The next window asks for the name and the directory where you would like to save the file:



After specifying your choice for the name and directory, click on **"Save"** to save the active file.

Close a file

To close the active file, click on **"File"** and then select **"Close"**.

Exit the software

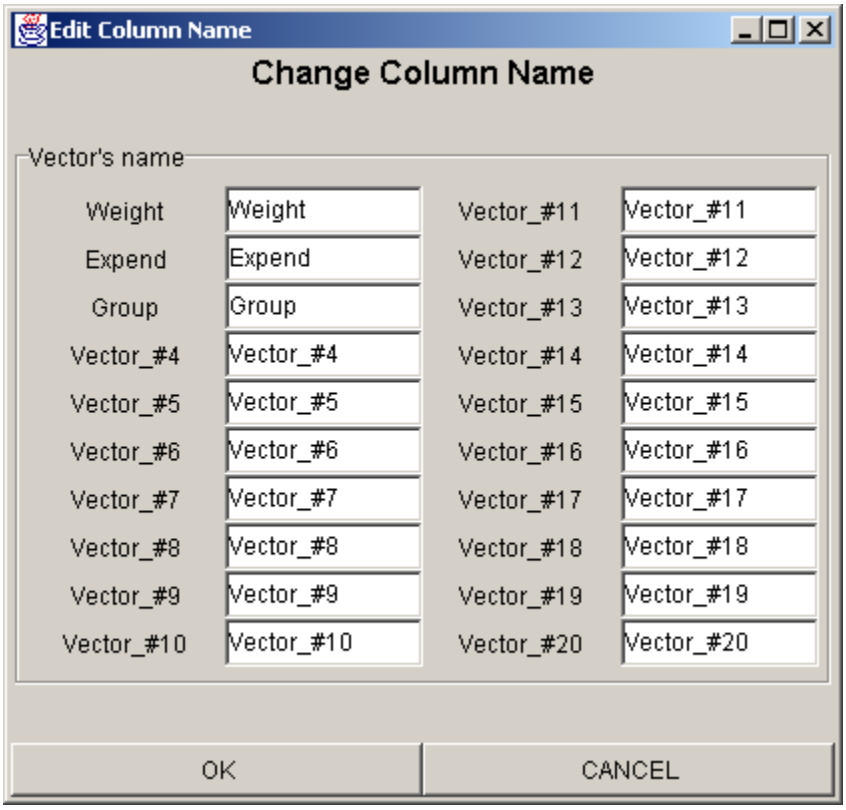
To exit the software, click on **"File"** and then select **"Exit"**.

Modifying the database

DAD offers the possibility to modify the dimension of a database and also to generate a new vector of data using logical or arithmetic operators.

Changing the names of vectors

To change the names of vectors, click on the button **"Edit"** and then select the item **"Change column name"**. The following windows appears:



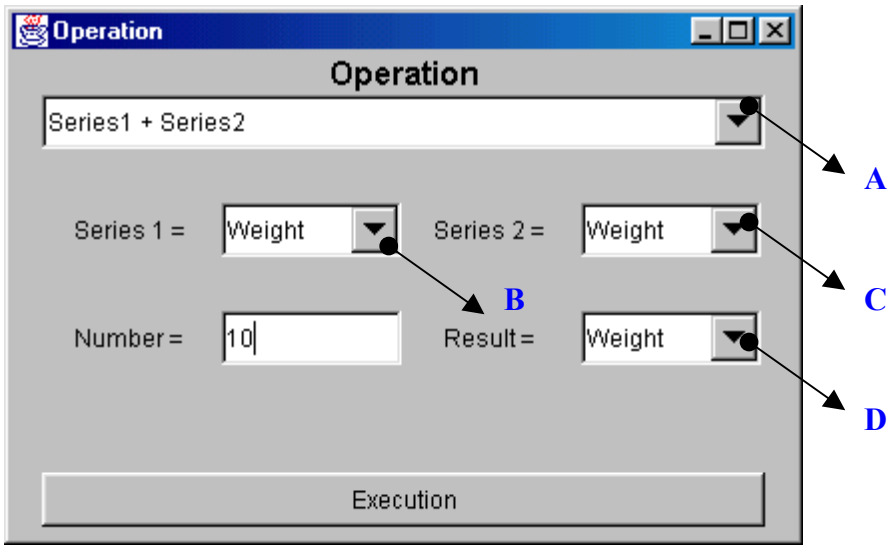
You can insert the new name of a vector and click on the button “OK” to confirm the change.

Generating new vectors

You may need to generate a new vector in the active database. The following steps describe the necessary procedures for this:

1- In the main menu, choose the command **"Edit"** and select the item **"Edition of columns".**

The next window appears for the specification of the type of operation that you wish to apply:



1- Choose the type of operation you need to carry out by clicking on the icon "**A**".
2- Select the vectors to be used to generate the new vector by clicking on the icons " **B**" and "**C**".
3- If a number is used to generate the new vector, write its value after "**Number".** By default, this number is set to 10.
4- Select the vector of results by clicking on the icon "**D**".

Denote vector 1 by S1(i) and vector 2 by S2(i). The following table then presents the type of operations available and their results.

Type of operation	Results
Series 1 + Series 2	**S1(i) + S2(i)**
Series 1 - Series 2	**S1(i) - S2(i)**
Series 1 * Series 2	**S1(i) * S2(i)**
Series 1 / Series 2	**S1(i) / S2(i)**
Series 1 + Number	**S1(i) + Number**
Series 1 - Number	**S1(i) - Number**
Series 1 * Number	**S1(i) * Number**
Series 1 / Number	**S1(i) / Number**
Exp (Series 1)	**Exp(S1(i))**
Log (Series 1)	**Log(S1(i))**
Series 1 = Series 2	**1 :if S1(i) = S2(i), otherwise 0**
Series 1 = Number	**1 :if S1(i) = S2(i), otherwise 0**
Series 1 ≥ Series 2	**1 :if S1(i) ≥ S2(i), otherwise 0**
Series 1 ≥ Number	**1 :if S1(i) ≥ S2(i), otherwise 0**
Series 1 ≤ Series 2	**1 :if S1(i) ≤ S2(i), otherwise 0**
Series 1 ≤ Number	**1 :if S1(i) ≤ S2(i), otherwise 0**

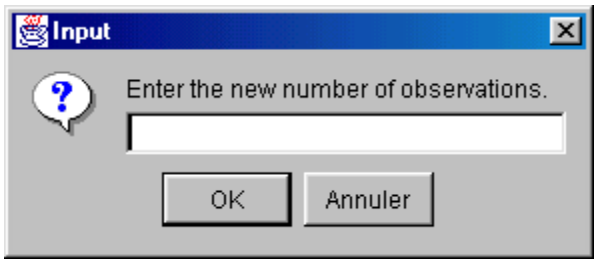
5- Finally, click on the button **"Execution"** to generate the new vector.

Copy, paste and clear commands

You can select some cells with your mouse and use the commands copy, paste, and clear to edit your database.

GetOBS and SetOBS commands

To obtain the number of observations of your active file, choose the command "GetOBS". If you would like to set a new number of observations, choose the command "SetOBS". The following window appears:



After this, enter the new number of observations and click on the button OK. The first SetOBS observations will now be used for the computations.

Changing the names of spreadsheet

To change the name of the spreadsheet, from the main menu, select the item "Edit⇒Change current sheet name" and indicate the new name.

Dimension of the spreadsheet

The length of the spreadsheet varies according to the following:

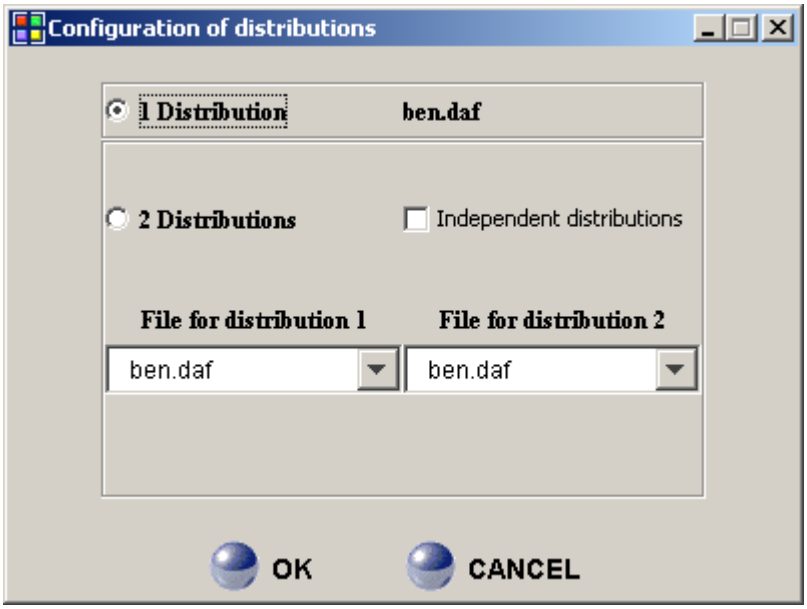
- By default, the length of the spreadsheet is 160 000 observations. This is done when a new file is created.
- If you download an ASCII file, the length of spreadsheet corresponds to the number of observations read from this file.
- In all cases, you can specify explicitly a desired length for the spreadsheet by indicating the new length after choosing the command "Edit" and the item "Enter the new length of the spreadsheet"

The new length of the spreadsheet cannot be below the number of observations OBS. The number of columns fixes the width of the spreadsheet. By default the number of columns is 16.

Applications in DAD

Introduction to applications

Remember that DAD can activate one or two databases. Once a database is activated, the user can then call different applications of DAD. Before you reach those applications, however, you must indicate how many databases are to be used in the application, and which ones. This is done through the following window:

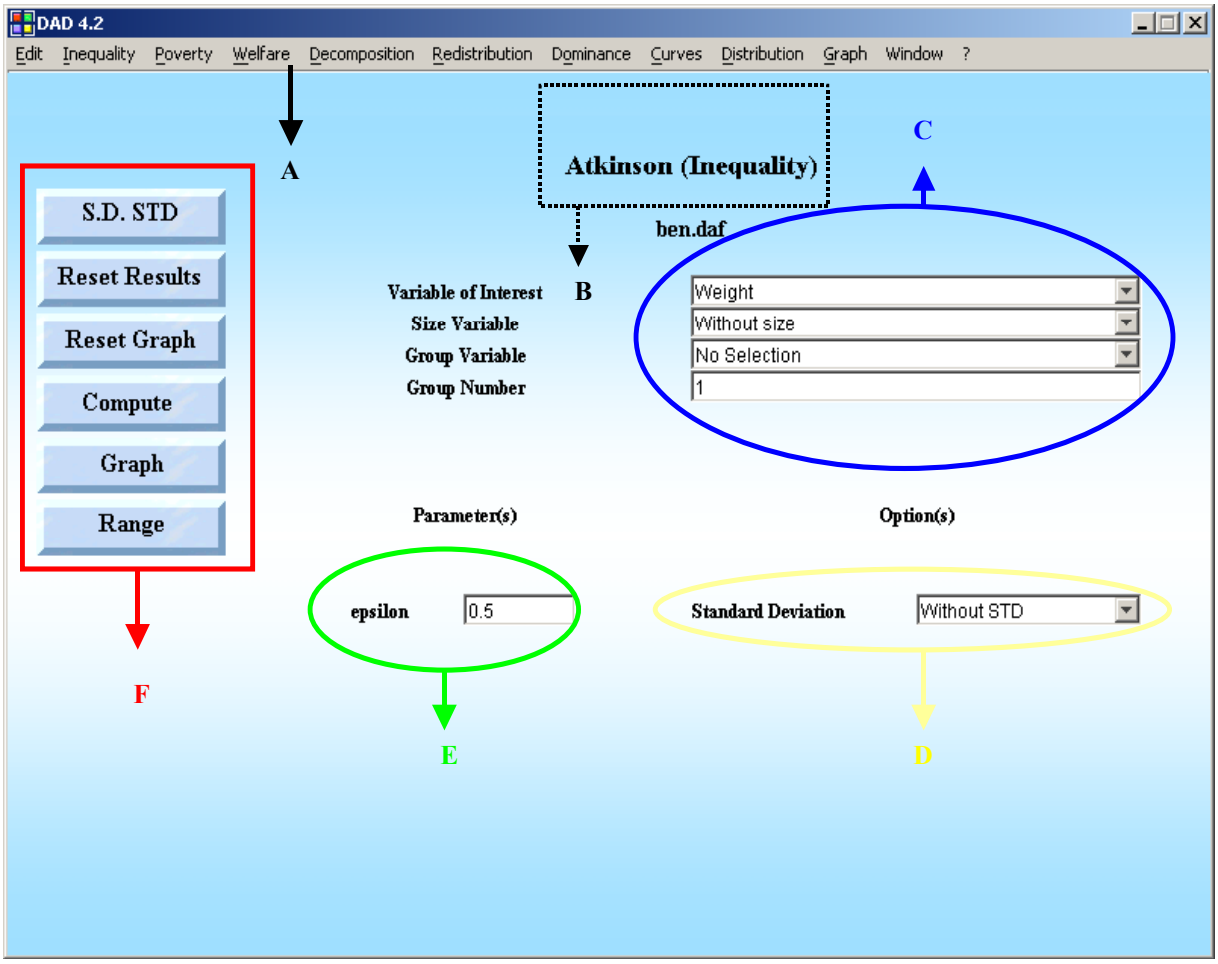


Each database represents one distribution. Generally, you should indicate the following information:

1- The number of distributions
2- The name of the file representing the first distribution.
3- The name of the file representing the second distribution.
4- When two distributions are to be used, you should indicate if the two distributions represent dependent or independent samples for the accurate computation of standard errors that use information on the joint distribution.

Confirm your choice by clicking on the button **"OK".** Once the choice is confirmed, you can reach the desired application.

Remark: If the number of distributions is one, the activated file is automatically the file specified on the 1st line.



A: Main menu
B: The name of the application and the name of the file used
C: Set of variables and parameters to be chosen as:

- Choice of variable of interest.
- Choice of size variable.
- Choice of group variable.
- Choice of group number.

D: Option to compute with or without standard deviation.
E: Parameters to be specified.
F: Set of Commands for this application.

You can to specify a weighting vector in order to weight your observations. Also, options shown in C allow you to compute an estimate for one specific group (or sub-sample) or sub-vector. The following example illustrates those different options.

Example

Suppose that you wish to compute the mean of a variable *y*, with y_i^j, denoting the i^{th} observation –household- of a person j. We call the vector to be used the "Variable of Interest". The following table displays the observations of *y* for a sample of ten households. The vector of sw_i **("Sampling Weight variable")** is the sampling weight to be applied to these observations and s_i is the size of observation -household- i. We can also assign to each of these observations a code c_i that indicates the subgroup of the population to which the i^{th} observation belongs. For example, code 1 may indicate that households live in town "V1" and code 2 that they live in town "V2":

Observation	y_i	c_i	sw_i	s_i
i	**Variable of interest**	**Group Variable**	**Sampling Weight variable**	**Size Variable**
1	500	1	3	2
2	200	2	1	1
3	300	1	1	4
4	1000	1	2	5
5	700	2	3	5
6	450	1	1	7
7	300	1	1	3
8	200	2	3	3
9	300	2	2	4
10	400	1	1	8

The user then has six possibilities for computing the mean, as shown in the following table:

	The mean	**Variable of Interest**	**Size Variable**	**Group Variable**	**Index of group**
1	For the 10 households **Without size**	y_i	Without Size	No selection	1 (*)
2	For the 10 households **With size**	y_i	s_i	No selection	1 (*)
3	For households living in town V1 **Without size**	y_i	Without Size	c_i	1
4	For households living in town V1 **With size**	y_i	s_i	c_i	1
5	For households living in town V2 **Without size**	y_i	Without Size	c_i	2
6	For households living in town V2 **With size**	y_i	s_i	c_i	2

1- (*): This choice does not affect the results since no group variable has been selected.
2- Consult the Sampling design section to know how can we initialise the sampling weight.

Finally, to compute the standard deviation on the estimate of the mean, you just need to select the option of computing **"with STD"**.

Basic Notation in DAD

In this following table, we present the basic notations used in the user manual of DAD.

Symbol	Indication
y	the variable of interest.
$\mathbf{y_i}$	the value of the variable of interest for observation i
sw	the Sampling Weight.
$\mathbf{sw_i}$	the Sampling Weight for observation i.
s	the size variable.
$\mathbf{s_i}$	the size of observation i (for example the size of household i).
$\mathbf{w_i}$	$sw_i * s_i$
c	the group variable.
$\mathbf{c_i}$	the group of observation i.
k	A group value (an integer).
$\mathbf{w_i^k}$	$\mathbf{w_i^k=1}$ if $\mathbf{c_i = k}$, and $\mathbf{w^k_i=0}$ otherwise.

Example: The mean of group k, $\mu(k)$, is then estimated as:

$$\mu(k) = \frac{\sum_{i=1}^{n} w_i^k y_i}{\sum_{i=1}^{n} w_i^k}$$

Taking into account sampling design in DAD

Sampling Design and DAD

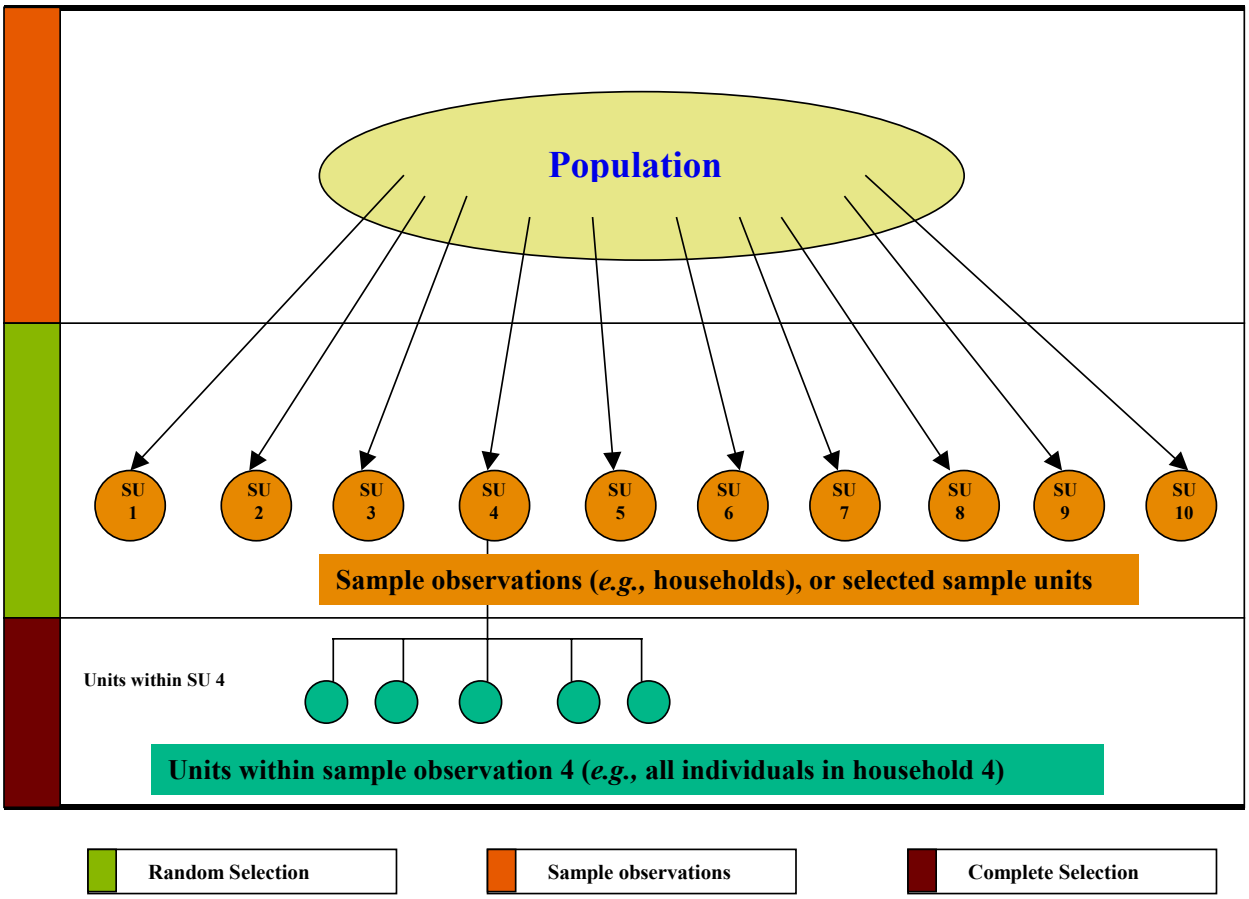
With version 4.2 and higher of DAD, the Sampling Design (SD) of the database can be specified in order to calculate the correct asymptotic sampling distribution of the various indices and statistics provided by DAD.

Data from sample surveys usually display four important characteristics:

1- they come with sampling weights (SW), also called inverse probability weights;
2- they are stratified;
3- they are clustered;
4- sample observations provide aggregate information (such as household expenditures) on a number of "statistical units" (such as individuals)

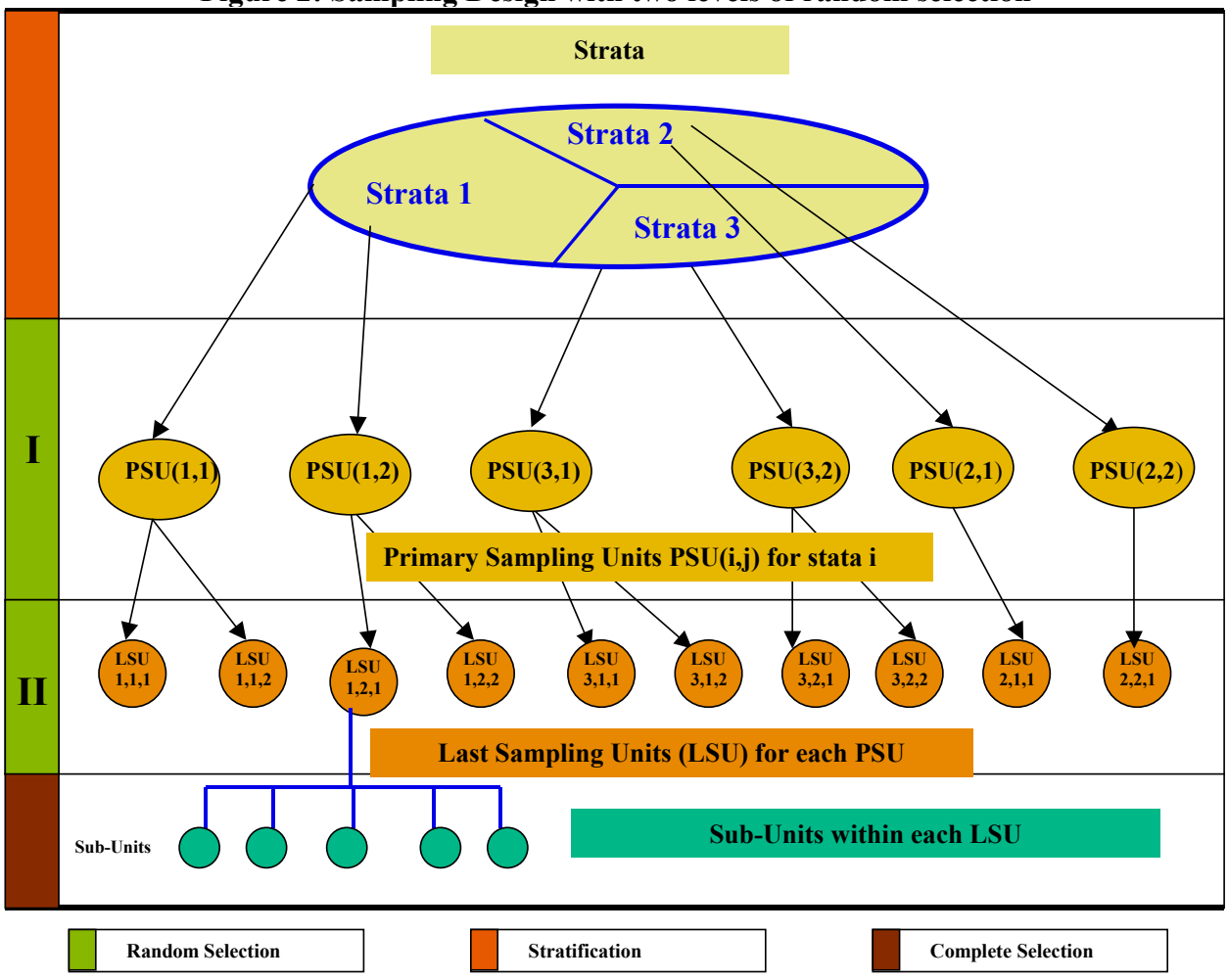
Figure 1 shows a graphical SD representation for the case of Simple Random Sampling (SRS), in which it is supposed that sample observations are directly and randomly selected from a base of sampling units (SUs) (*e.g.,* the list of all households within in a country).

Figure 1: Simple Random Sampling



SRS is rarely used to generate household surveys. Hence, most SD encountered in practice will not look like that in Figure 1. Most SD will look instead like that of Figure 2. A country is first divided into geographical or administrative zones and areas, called *strata*. Each zone or area thus represents a strata in Figure 2. The first random selection takes place within the Primary Sampling Units (denoted as PSU's) of each stratum. Within each stratum, a number of PSU's are randomly selected. This random selection of PSU's provides "clusters" of information. PSU's are often provinces, departments, villages, *etc* Within each PSU, there may then be other levels of random selection. For instance, within each province, a number of villages may be randomly selected, and within every selected village, a number of households may be randomly selected. The final sample observations constitute the Last Sampling Units (LSU's). Each sample observation may then provide aggregate information (such as household expenditures) on all individuals or agents found within that LSU. These individuals or agents are *not* selected – information on all on them appears in the sample. They therefore do not represent the LSUs in statistical terminology.

Figure 2: Sampling Design with two levels of random selection



Impact of SD on the sampling error of DAD's estimators

a) Impact of stratification

Generally speaking, a variable of interest, such as household income, tends to be less variable within strata than across the entire population. This is because households within the same stratum typically share to a greater extent than in the entire population some socio-economic characteristics, such as geographical locations, climatic conditions, and demographic characteristics,and that these characteristics are determinants of the living standards of these households. Stratification ensures that a certain number of observations are selected from each of a certain number of strata. Hence, it helps generate sample information from a diversity of "socio-economic areas". Because information from a "broader" spectrum of the population leads on average to more precise estimates, stratification generally decreases the sampling variance of estimators. For instance, suppose at the extreme that household income is the same for all households in a stratum, and this, for all strata. In this case, supposing also that the population size of each stratum is known, it is sufficient to draw one household from each stratum to know exactly the distribution of income in the population.

b) Impact of clustering (or multi-stage sampling)

Multi-stage sampling implies observations end up in a sample only subsequently to a process of multiple selection. "Groups" of observations are first randomly selected within a population (which may be stratified); this is followed by further sampling within the selected groups, which may be followed by yet another process of random selection within the subgroups selected in the previous stage. The first selection stage takes place at the level of PSU's, and generates what are often called "clusters". Generally, variables of interest (such as living standards) vary less within a cluster than between clusters. Hence, multi-stage selection reduces the "diversity" of information generated by sampling. The impact of clustering sample observations is therefore to tend to decrease the precision of populations estimators, and thus to increase their sampling variance. *Ceteris paribus*, the lower the variability of a variable of interest within clusters, the larger the loss of information that there is in sampling further within the same clusters. To see this, suppose for instance an extreme case in which household income happens to be the same for all households in a cluster, and this, for all clusters. In such cases, it is clearly wasteful to adopt multi-stage sampling: it would be sufficient to draw one household from each cluster in order to know the distribution of income within that cluster. It would be more informative to draw randomly other clusters.

Sampling Design in DAD

By default, when a data file is loaded in DAD, the type of SD assigned to the data is the SRS presented in Figure 1. Once the data are loaded, the exact SD structure can nevertheless be easily specified. Up to 5 vectors can help specify that structure:

Table 1: Description of vectors used in DAD to specify the SD

Vectors	Description
Strata	Specifies the name of the variable (integer type) that contains stratum identifiers
PSU	Specifies the name of the variable (integer type) that contains identifiers for the Primary Sampling Units
LSU	Specifies the name of the variable (integer type) that contains identifiers for the Last Sampling Units
SW	Specifies the name of the variable for the Sampling Weights. Sampling weights are the inverse of the sampling rate. Roughly speaking, they equal the number of observations in the underlying population that are represented by each sample observation.
FPC	Specifies the name of the variable for the Finite Population Correction factor.
	With FPC, DAD derives an indicator f_h for each observation h, which is then used to compute SD-corrected sampling errors. If the variable FCP is not specified, f_h=0 for all observations; When the variable specified has values <= 1, it is directly interpreted as a stratum sampling rate f_h =n_h/N_h, where n_h = number of PSUs sampled from the strata to which h belongs and N_h = total number of PSUs in the population belonging to stratum h. When the variable specified has values greater than or equal to n_h, it is interpreted as representing N_h; f_h is then set to n_h/N_h.

The following table contains an example of vectors used to specify the type of SD shown in Figure 2.

Table 2: Example of SD.

OBS	Strata	PSU	LSU	SW
1	1	1	1	6
2	1	1	2	6
3	1	2	1	6
4	1	2	2	6
5	3	1	1	5
6	3	1	2	5
7	3	2	1	5
8	3	2	2	5
9	2	1	1	3
10	2	2	1	3
SUM	**3**	**6**	**10**	**50**

Omitting SW will systematically bias both the estimators of the values of indices and points on curves as well as the estimation of the sampling variance of those estimators. Consider for instance the estimation of total population income from the data shown in table 2. 4 households appear in strata 1, but the population number of households in that strata is six times as large (that

is, 24), and this is captured by the SW variable. Total population income for strata 1 would therefore be estimated to be six times that of total sample income for strata 1.

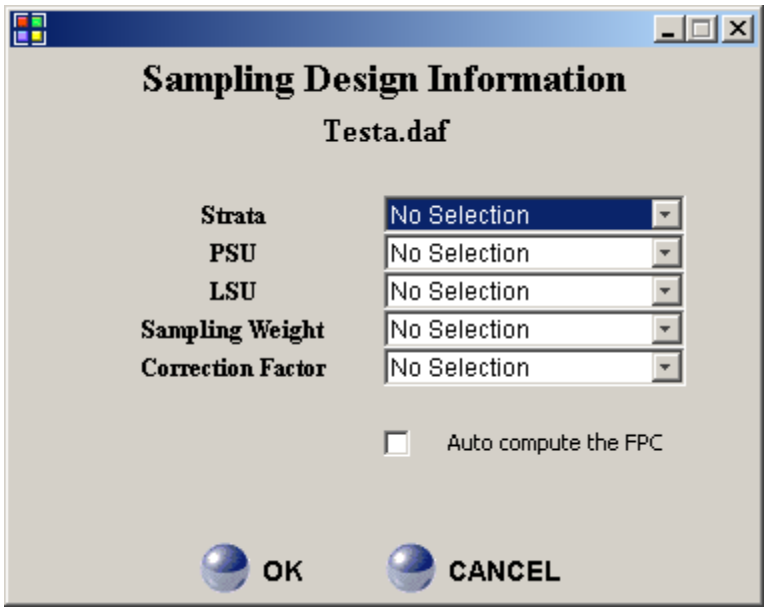
Table 3: Example of SD.

OBS	Strata	LSU	SW	N_h
1	1	1	6	24
2	1	2	6	24
3	1	3	6	24
4	1	4	6	24
5	3	1	5	20
6	3	2	5	20
7	3	3	5	20
8	3	4	5	20
9	2	1	3	6
10	2	2	3	6
SUM	**3**	**10**	**50**	**---**

The FPC factor accounts for the reduction in sampling variance that occurs when a sample is drawn *without* replacement from a finite population (as compared to sampling *with* replacement). According to table 3, the four LSU's of strata 1 were selected without replacement from a population of 24 LSU's. These fuor LSU's are then necessarily distinct by design. If sampling had been done with replacement, then multiple observations of the same population LSU's could have been generated. Because sampling without replacement guarantees that sample observations represent different sampling units, it therefore generates greater sampling information and leads to smaller sampling variances than with sampling with replacement. For strata 1 of Table 3, data from four distinct LSU's (or PSU's) out of 24 are necessarily generated after sampling. The f_h factor for that strata is then 4/24=0.1666.

Important Remark: *We can initialise and use the FPC correction just when the SD is based on one stage of random selection of LSU's. In this case PSU's and LSU's are equivalent.*

To initialize the SD after loading the database, select from the main menu the item "Edit->Set Sample Design". The following window then appears.



This allows DAD to take into account a wide variety of possible SD. This is made by selecting (or not selecting) vectors for any of the five choices offered above. In the case of SRS within a number of strata, there would be an indicator of a strata vector without any indication of a vector of PSU's. The following table presents some of these combinations.

Strata	PSU	LSU	SW	FPC	Indication
					SD is SRS without sampling weights
X			X		SD is stratified with SW
	X	X	X		No stratification, but multi-stage sampling and SW
		X	X		Random (one-stage) sampling of LSU's with LSU-specific selection probabilities. This can occur for instance if, once an individual is selected, all individuals in his household are also automatically selected. Implicitly, then, it is the household that is selected as a LSU
X	X		X		Stratification with only the first sampling stage specified by the user
X		X			Stratification with one-stage sampling and sampling weights (wrongly?) omitted
X		X	X		Stratification with one-stage sampling and sampling weights (wrongly?) omitted
X	X	X			Stratification with multi-stage sampling and sampling weights (wrongly?) omitted
X	X	X	X		Stratification with multi-stage sampling and sampling weights provided
X	X	X	X	X	Stratification with multi-stage sampling and sampling weights provided. The finite population correction factor is also provided; this supposes that sampling for the statistical inferences

X: Indicate that the variable is selected

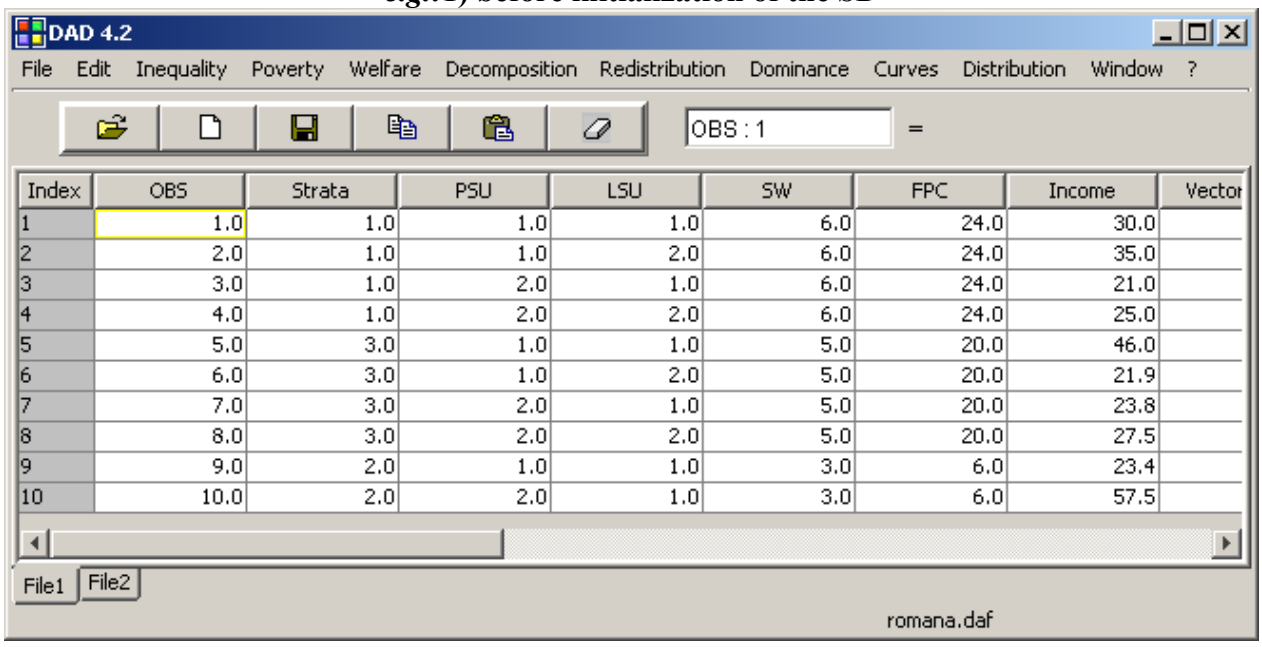
Note that when DAD finds the values of the strata-psu-lsu variables to be the same across observations, it supposes that these observations comefrom just one LSU.

If the option "Auto-compute FPC" is activated, DAD generates implicitly the FPC vector.

Remarks:

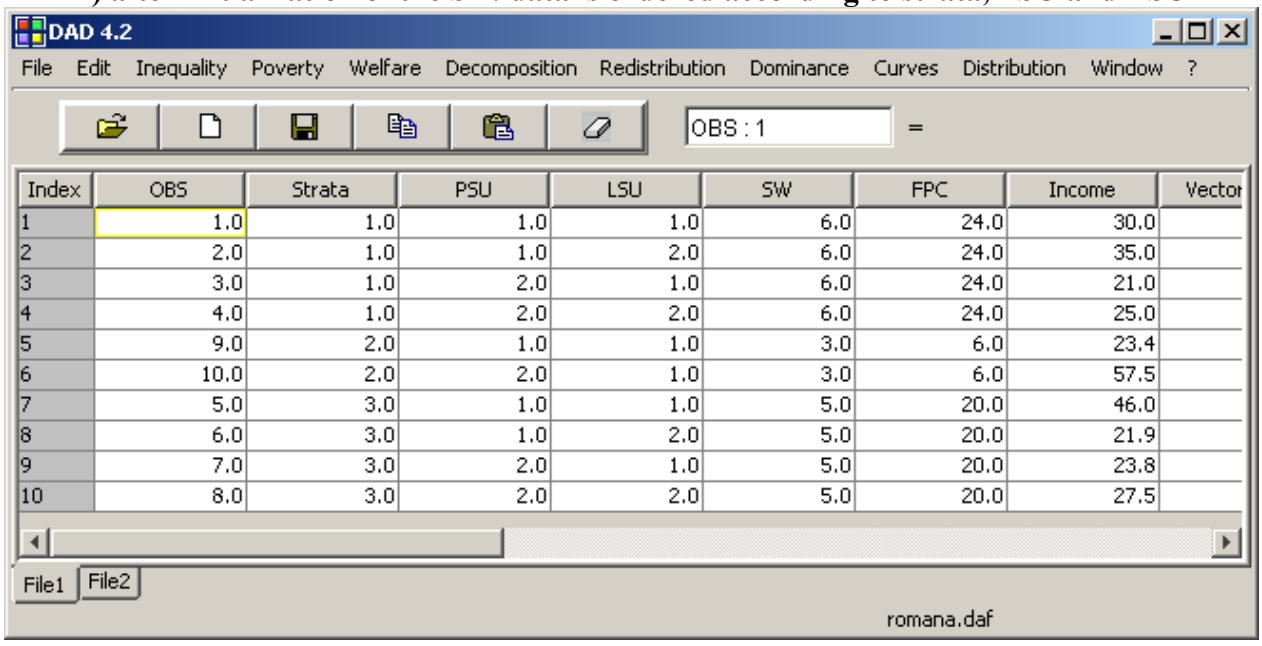
- *After initialization of the SD information, the dataset is automatically ordered by (when specified) strata, PSU's and LSU's.*
- *There should be more than one PSU within each stratum.*

***e.g.*:1) before initialization of the SD**



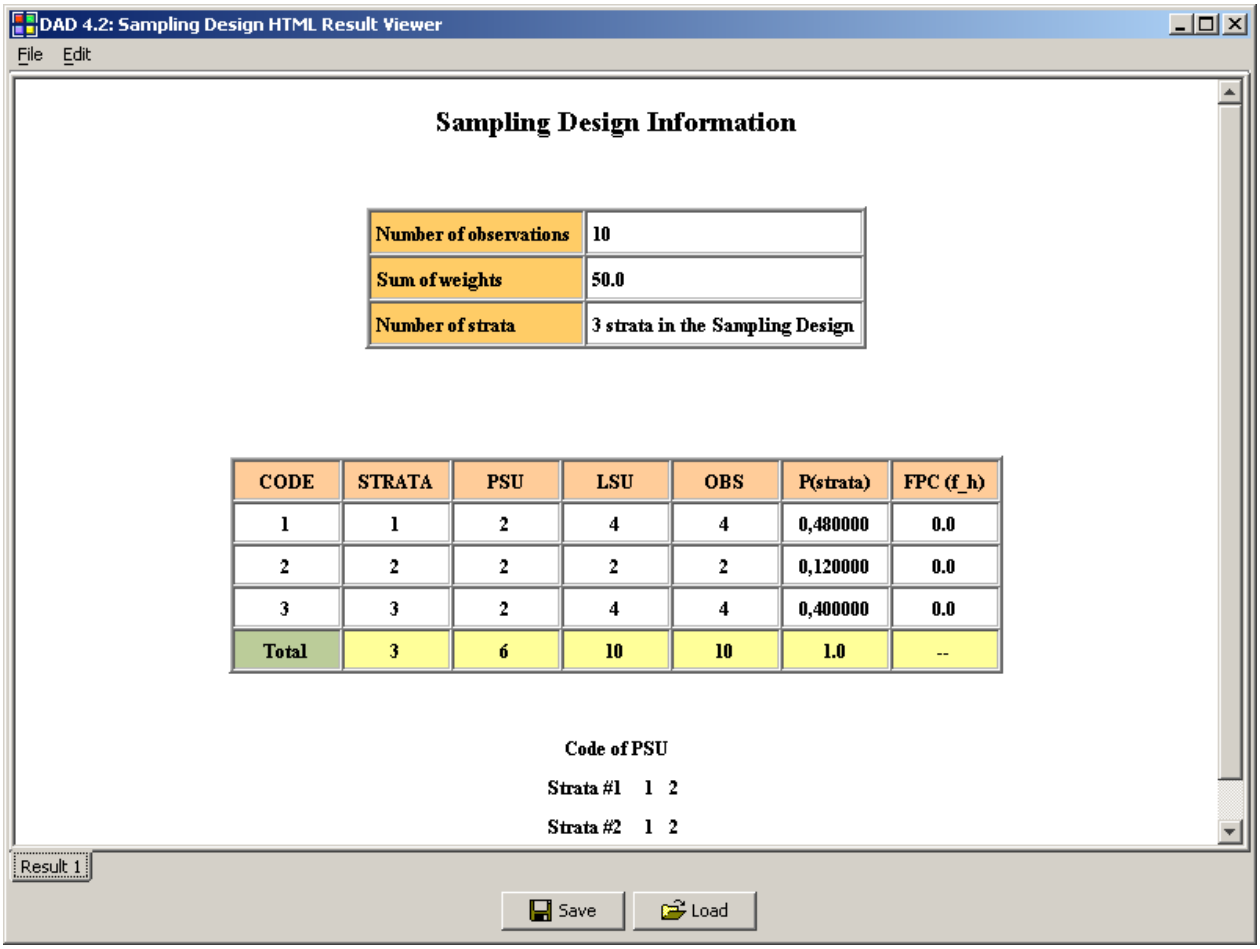
Index	OBS	Strata	PSU	LSU	SW	FPC	Income	Vector
1	1.0	1.0	1.0	1.0	6.0	24.0	30.0	
2	2.0	1.0	1.0	2.0	6.0	24.0	35.0	
3	3.0	1.0	2.0	1.0	6.0	24.0	21.0	
4	4.0	1.0	2.0	2.0	6.0	24.0	25.0	
5	5.0	3.0	1.0	1.0	5.0	20.0	46.0	
6	6.0	3.0	1.0	2.0	5.0	20.0	21.9	
7	7.0	3.0	2.0	1.0	5.0	20.0	23.8	
8	8.0	3.0	2.0	2.0	5.0	20.0	27.5	
9	9.0	2.0	1.0	1.0	3.0	6.0	23.4	
10	10.0	2.0	2.0	1.0	3.0	6.0	57.5	

2) after initialization of the SD: data is ordered according to strata, PSU and LSU



Index	OBS	Strata	PSU	LSU	SW	FPC	Income	Vector
1	1.0	1.0	1.0	1.0	6.0	24.0	30.0	
2	2.0	1.0	1.0	2.0	6.0	24.0	35.0	
3	3.0	1.0	2.0	1.0	6.0	24.0	21.0	
4	4.0	1.0	2.0	2.0	6.0	24.0	25.0	
5	9.0	2.0	1.0	1.0	3.0	6.0	23.4	
6	10.0	2.0	2.0	1.0	3.0	6.0	57.5	
7	5.0	3.0	1.0	1.0	5.0	20.0	46.0	
8	6.0	3.0	1.0	2.0	5.0	20.0	21.9	
9	7.0	3.0	2.0	1.0	5.0	20.0	23.8	
10	8.0	3.0	2.0	2.0	5.0	20.0	27.5	

To show the SD information, select from main menu the item “Edit->Summarize Sample Design”. The following window appears.



Computation of standard errors in DAD

This section shows how the standard errors of DAD's estimators of distributive indices and curves are computed. The methodology is based on the asymptotic sampling distribution of such indices and curves. All of DAD's estimators are asymptotically normally distributed around their true population value. As will be discussed below, we expect this methodology to provide a good approximation to the true sampling distribution of DAD's estimators for relative large samples.

Estimators of the distributive indices

Estimators of distributive indices (such as poverty and inequality indices) take the following general form:

$$\hat{\theta} = g(\hat{\alpha}_1, \hat{\alpha}_2, \cdots \hat{\alpha}_K) \text{ with } \alpha_k \text{ asymptotically expressible as } \alpha_k = \sum_{j=1}^{m} y_{k,j}$$

where θ can be expressed as a continuous function g of the α's, m is the number of sample observations and $y_{k,j}$ is usually some transform of the living standard of individual or household j. We use Rao's (1973) linearization approach [1] to derive the standard error of these distributive indices. This approach says that the sampling variance $\hat{\theta}$ equals the variance of a linear approximation of $\hat{\theta}$:

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}\left(\frac{\partial \theta}{\partial \alpha_1}(\hat{\alpha}_1 - \alpha_1) + \frac{\partial \theta}{\partial \alpha_2}(\hat{\alpha}_2 - \alpha_2) + \cdots + \frac{\partial \theta}{\partial \alpha_K}(\hat{\alpha}_K - \alpha_K) \right)$$

In matrix format, the variance of $\hat{\theta}$ is given by

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(V'MV)$$

with M the covariance matrix of the $\hat{\alpha}$ and V the gradient of θ:

[1] Rao,C.R. (1973). *Linear Statistical Inference and Its Application*. New York: Wiley.

$$V = \begin{pmatrix} \dfrac{\partial \theta}{\partial \alpha_1} \\ \dfrac{\partial \theta}{\partial \alpha_2} \\ \vdots \\ \dfrac{\partial \theta}{\partial \alpha_K} \end{pmatrix}$$

The gradient elements $\left(\frac{\partial \theta}{\partial \alpha_1}, \frac{\partial \theta}{\partial \alpha_2}, \cdots \right)$ can be estimated consistently using estimates $\left(\frac{\partial \hat{\theta}}{\partial \hat{\alpha}_1}, \frac{\partial \hat{\theta}}{\partial \hat{\alpha}_2}, \cdots \right)$ of the true derivatives. The covariance matrix is defined as

$$M = \begin{vmatrix} \mathrm{Var}(\alpha_1) & \mathrm{Cov}(\alpha_1, \alpha_2) & & \mathrm{Cov}(\alpha_1, \alpha_K) \\ \mathrm{Cov}(\alpha_2, \alpha_1) & \mathrm{Var}(\alpha_2) & \cdots & \mathrm{Cov}(\alpha_2, \alpha_K) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\alpha_K, \alpha_1) & \mathrm{Cov}(\alpha_K, \alpha_2) & \cdots & \mathrm{Var}(\alpha_K) \end{vmatrix}$$

The elements of the covariance matrix are again estimated consistently using the sample data, replacing for instance $\mathrm{Var}(\hat{\alpha})$ by $\hat{\mathrm{V}}\mathrm{ar}(\hat{\alpha})$. It is at the level of the estimation of these covariance elements that the full sampling design structure is taken into account.

Finite-sample properties of asymptotic results

It may be instructive to compare the results of the above asymptotic approach to those of a numerical simulation approach like the bootstrap. The bootstrap (BTS) is a method for estimating the sampling distribution of an estimator which proceeds by re-sampling repetitively one's data. For each simulated sample, one recalculates the value of this estimator and then uses that BTS distribution to carry out statistical inference. In finite samples, neither the asymptotic nor the BTS sampling distribution is necessarily superior to the other. In infinite samples, they are usually equivalent.

Bootstrap and simple random sampling

The following steps the BTS approach for a sample drawn using Simple Random Sampling:

1- Draw *with replacement* m observations from the initial sample.
2- Compute the distributive estimator from this new generated sample.
3- Repeat the first two steps N times.
4- Compute the variance or the BTS distributions using these N generated estimators.

Bootstrap and complex sampling design

The steps here are similar to those above with Simple Random Sampling. Only the first step differs to take into account the precise way in which the original sample was drawn. Suppose for example that:

- The data were drawn from two strata, with m1 observations in stratum 1 and m2 observations in stratum 2.
- Observations in every stratum were selected randomly with equal probabilities

The first step will then consist in selecting randomly and with the same probability m1 observations from stratum1 and (independently) m2 observations from stratum2. Aggregating these two sub samples will yield the new generated sample. Repeating this N times will generate the BTS sampling distribution.

Illustrations

The following table presents the sampling design information of a hypothetical sample of 800 observations.

Sampling Design Information

Number of observations	**800**
Sum of weights	**6200.0**
Number of strata	**2 strata in the Sampling Design**

CODE	STRATA	PSU	LSU	OBS	P(strata)	FPC (f_h)
1	1	30	300	300	0,193548	0.0
2	2	50	500	500	0,806452	0.0
Total	2	80	800	800	1.0	--

The following tables present estimates of the standard errors of some distributive indices using asymptotic theory (DAD) and the BTS procedure.

Atkinson Index (ε =0.5) = 0,09131119						
W	Strata	Psu	Lsu	Size =psu	St.err. DAD	St.err. BTS
×					0,00403011	0,00404464
×	×				0,00396117	0,00391402
×		×			0,00479089	0,00473645
×	×	×			0,00414549	0,00412479
×	×	×		×	0,00455368	0,00454454

FGT (α =1; z=3000) = 566.47774194						
W	Strata	Psu	Lsu	Size =psu	St.err. DAD	St.err. BTS
×					30,15130207	30,31106186
×	×				29,76615787	29,82831383
×		×			34,90968660	34,49846649
×	×	×			31,21606735	31,36449814
×	×	×		×	40,20904414	40,10400009

Lorenz (p=0.5) =0,26371264						
W	Strata	Psu	Lsu	Size =psu	St.err. DAD	St.err. BTS
✕					0,00618343	0,00617247
✕	✕				0,00612036	0,00614563
✕		✕			0,00695073	0,00697490
✕	✕	✕			0,00632417	0,00636899
✕	✕	✕		✕	0,00726710	0,00724934

Gini (ρ =2) = 0,42403734						
W	Strata	Psu	Lsu	Size =psu	St.err. DAD	St.err. BTS
✕					0,00801557	0,00809321
✕	✕				0,00786047	0,00781983
✕		✕			0,00964692	0,00964823
✕	✕	✕			0,00820847	0,00827642
✕	✕	✕		✕	0,00949502	0,00946204

Notes:

W	Sampling weight
✕	Sampling-design feature is used

Inequality

y_i is the living standard of observation i. We assume that the n observations have been ordered in increasing values of y, such that $y_i \le y_{i+1}, \forall i = 1,...,n-1$.

The Atkinson index

Denote the Atkinson index of inequality for the group k by $I(k;\varepsilon)$. It can be expressed as follows:

$$I(k;\varepsilon) = \frac{\mu(k) - \xi(k;\varepsilon)}{\mu(k)} \quad \text{where} \quad \mu(k) = \frac{\sum_{i=1}^{n} w_i^k y_i}{\sum_{i=1}^{n} w_i^k}$$

The Atkinson index of social welfare is as follows:

$$\xi(k;\varepsilon) = \begin{cases} \left[\frac{1}{\sum_{i=1}^{n} w_i^k} \sum_{i=1}^{n} w_i^k (y_i)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} & \rightarrow \text{if } \varepsilon \neq 1 \quad \text{and } \varepsilon \geq 0 \\ \text{Exp}\left[\frac{1}{\sum_{i=1}^{n} w_i^k} \sum_{i=1}^{n} w_i^k \ln(y_i) \right] & \rightarrow \varepsilon = 1 \end{cases}$$

Case 1 : One distribution

If you wish to compute the Atkinson index of inequality for only one distribution, follow these steps:

1- From the main menu, choose "Inequality⇒ Atkinson index".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and values of parameters as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
epsilon	ε	Compulsory

Among the buttons, you find the following commands:

- "**Compute":** to compute the Atkinson index. If you also want the standard deviation of this index, choose the option for computing with a standard deviation.
- "**Graph":** to draw the value of the index according to the parameter ε. If you want to specify a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.

Case 2: Two distributions

To compute the Atkinson index of two distributions:

1- From the main menu, choose the item: "Inequality⇒ Atkinson index".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_1	Optional
epsilon	ε_1	ε_2	Compulsory

Among the buttons, you find the command « **Compute ».** To compute the standard deviation of this index, choose the option for computing with standard deviation.

S-Gini index

Denoting the S-Gini index of inequality for the group k by $I(k;\rho)$, and the S-Gini social welfare index by $\xi(k;\rho)$, we have:

$$I(k;\rho) = \frac{\mu(k) - \xi(k;\rho)}{\mu(k)}$$

where

$$\xi(k;\rho) = \sum_{i=1}^{n} \left[\frac{(V_i)^\rho - (V_{i+1})^\rho}{[V_1]^\rho} \right] y_i$$

and

$$V_i = \sum_{h=i}^{n} w_h^k$$

Case 1: One distribution

To compute the S-Gini index of inequality for only one distribution:

1- From the main menu, choose the item: "Inequality⇒ S-Gini index".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and values of parameters as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
rho	ρ	Compulsory

Two choices of commands appear among the buttons:

- **"Compute":** to compute the S-Gini index. To compute the standard deviation of this index, choose the option for computing with standard deviation.
- **"Graph" :** to draw the value of the index according to the parameter ρ. To specify such a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.

Case 2: Two distributions

To reach the S-Gini application with two distributions:

1- From the main menu, choose the item: "Inequality⇒ S-Gini index".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	**Vectors or parameters**		**Choice is:**
	Distribution 1	**Distribution 2**	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
rho	ρ_1	ρ_2	Compulsory

Among the buttons, you will find the command « **Compute ».** To compute the standard deviation of this index, choose the option for computing with standard deviation.

The Atkinson-Gini index

Denoting the Atkinson-Gini index of inequality for the group k by $I(k;\varepsilon,\rho)$, and the S-Gini social welfare index by $\xi(k;\varepsilon,\rho)$, we have:

$$I(k;\varepsilon,\rho)=\frac{\mu(k)-\xi(k;\varepsilon,\rho)}{\mu(k)}$$

where

$$\xi(k;\varepsilon,\rho)=\begin{cases}\left[\sum_{i=1}^{n}\left[\frac{(V_i)^{\rho}-(V_{i+1})^{\rho}}{(V_1)^{\rho}}\right](y_i)^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} & \rightarrow \varepsilon\neq 1,\varepsilon\geq 0 \quad \text{and} \quad \rho\geq 1\\ \text{Exp}\left[\sum_{i=1}^{n}\left[\frac{(V_i)^{\rho}-(V_{i+1})^{\rho}}{(V_1)^{\rho}}\right]\ln(y_i)\right] & \rightarrow \varepsilon=1 \quad \text{and} \quad \rho\geq 1\end{cases}$$

and

$$V_i = \sum_{h=i}^{n} w_h^k$$

Case 1: One distribution

To compute this index of inequality for only one distribution:

1- From the main menu, choose the item: "Inequality⇒ Atkinson-Gini index".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	**Variables or parameters**	**Choice is:**
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
epsilon	ε	Compulsory
rho	ρ	Compulsory

Among the buttons you will find the command "**Compute",** which computes the Atkinson-Gini index. To compute the standard deviation of this index, choose the option for computing with standard deviation.

Case 2 : Two distributions

To reach the Atkinson-Gini application with two distributions:

1- From the main menu, choose the item: "Inequality⇒ Atkinson-Gini".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	**Distribution 2**	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
rho	ρ_1	ρ_2	Compulsory
epsilon	ε_1	ε_2	Compulsory

Among the buttons you will find the command « **Compute ».** To compute the standard deviation of this index, choose the option for computing with standard deviation.

The Generalised Entropy index of inequality

The Generalised Entropy Index of inequality for the group k is as follows:

$$I(k;\theta)=\begin{cases}\dfrac{1}{\theta(\theta-1)\sum_{i=1}^{n} w_i^k}\sum_i w_i^k\left[\left(\dfrac{y_i}{\mu(k)}\right)^{\theta}-1\right] & \text{if} \quad \theta\neq 0,1\\ \dfrac{1}{\sum_{i=1}^{n} w_i^k}\sum_i w_i^k \log\left(\dfrac{\mu(k)}{y_i}\right) & \text{if} \quad \theta=0\\ \dfrac{1}{\sum_{i=1}^{n} w_i^k}\sum_i \dfrac{w_i^k y_i}{\mu(k)}\log\left(\dfrac{y_i}{\mu(k)}\right) & \text{if} \quad \theta=1\end{cases}$$

Case 1 : One distribution

To compute the Generalised Entropy index of inequality for only one distribution:

1- From the main menu, choose the item: "Inequality⇒ Entropy index".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
theta	θ	Compulsory

Among the buttons., you find the following choices:

- "**Compute":** computes the Generalised Entropy index. To compute the standard deviation of this index, choose the option for computing with the standard deviation.
- "**Graph":** to draw the value of index according to the parameter θ. To specify a range for the horizontal axis, choose the item " Graph Management $\Rightarrow$ Change range of x " from the main menu.

Case 2 : Two distributions

To calculate the Generalised Entropy index for two distributions:

1- From the main menu, choose the item: "Inequality$\Rightarrow$ Entropy index".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	w^1	w^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
theta	θ_1	θ_2	Compulsory

Among the buttons, you will find the command « Compute ». To compute the standard deviation of this index, choose the option for computing with standard deviation.

The Quantile Ratio and the Interquantile Ratio Index

Denote the Quantile Ratio for group k by $QR(k;p_1,p_2)$; it can be expressed as follows:

$$QR(k;p_1,p_2)=\frac{Q(k,p_1)}{Q(k,p_2)}$$

where Q(k,p) denote the p-quantile of group k.

The Interquantile Ratio $IQR(k;p_1,p_2)$ is defined as:

$$IQR(k;p_1,p_2)=\frac{Q(k,p_1)-Q(k,p_2)}{\mu}$$

Remark: The instructions for the Interquantile Ratio are similar to those for the Quantile Ratio.

Case 1 : One distribution

If you wish to compute the Quantile Ratio for only one distribution, follow these steps:

1- From the main menu, choose "Inequality⇒ Quantile Ratio index".
2- In the configuration of the application, choose 1 distribution.
3- After confirming your choice, the application appears. Choose the different vectors and values of parameters as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Percentile for numerator	p_1	Compulsory
Percentile for denominator	p_2	Compulsory

Among the buttons., you will find the following command:

- "**Compute":** to compute the Quantile ration. If you also want the standard deviation on the estimator of that index, choose the option for computing with a standard deviation.

Case 2: Two distributions

To compute the Quantile Ratio index with two distributions:

1- From the main menu, choose the item: "Inequality⇒ Quantile Ratio index".
2- In the configuration of application, choose 2 as the number of distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	**Distribution 2**	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_1	Optional
Percentile for numerator	p_1	p_1	Compulsory
Percentile for denominator	p_2	p_2	Compulsory

Among the buttons, you will find the command « **Compute ».** To compute the standard deviation of the estimator of that index, choose the option for computing with standard deviation.

The Coefficient of Variation Index

Denote the Coefficient of Variation index of inequality for the group k by CV. It can be expressed as follows:

$$CV = \left[\frac{\sum_{i=1}^{n} w_i^k y_i^2 / \sum_{i=1}^{n} w_i^k - \mu^2}{\mu^2} \right]^{\frac{1}{2}}$$

Case 1: One distribution

If you wish to compute the Coefficient of Variation index of inequality for only one distribution, follow these steps:

1- From the main menu, choose the item "Inequality⇒ Coefficient of Variation ".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and values of parameters as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional

Among the buttons, you will find the following command:

- "**Compute":** to compute the Variation Logarithms index. If you also want the standard deviation of this index, choose the option for computing with a standard deviation.

Case 2: Two distributions

To compute the Coefficient of Variation of two distributions:

1- From the main menu, choose the item: "Inequality⇒ Coefficient of Variation ".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_1	Optional

Among the buttons, you will find the command « **Compute ».** To compute the standard deviation of this index, choose the option for computing with standard deviation.

The Logarithmic Variance Index

Denote the Logarithmic Variance index of inequality for the group k by LV; it can be expressed as follows:

$$LV = \frac{\sum_{i=1}^{n} w_i^k \left(\log(y_i) - lmu\right)^2}{\sum_{i=1}^{n} w_i^k} \quad \text{where} \quad lmu = \log\left(\frac{\sum_{i=1}^{n} w_i^k y_i}{\sum_{i=1}^{n} w_i^k}\right)$$

Case 1: One distribution

If you wish to compute the Logarithmic Variance index of inequality for only one distribution, follow these steps:

1- From the main menu, choose the following items "Inequality⇒ Logarithmic Variance ".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and values of parameters as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional

Among the buttons, you find the following command:

- "**Compute":** to compute the Logarithmic Variance index. If you also want the standard deviation of this index, choose the option for computing with a standard deviation.

Case 2: Two distributions

To compute the Logarithmic Variance index of two distributions:

1- From the main menu, choose the item: "Inequality⇒ Logarithmic Variance ".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_1	Optional

Among the buttons, you find the command « **Compute ».** To compute the standard deviation of this index, choose the option for computing with standard deviation.

The Variance of Logarithms

Denote the Variance of Logarithms index of inequality for group k by VL. It can be expressed as follows:

$$VL = \frac{\sum_{i=1}^{n} w_i^k \left(\log(y_i) - lmu\right)^2}{\sum_{i=1}^{n} w_i^k} \quad \text{where} \quad lmu = \frac{\sum_{i=1}^{n} w_i^k \log(y_i)}{\sum_{i=1}^{n} w_i^k}$$

Case 1 : One distribution

If you wish to compute the Variance of Logarithms index of inequality for only one distribution, follow these steps:

1- From the main menu, choose the item "Inequality⇒ Variance of Logarithms ".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and values of parameters as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional

Among the buttons, you will find the command:

- "**Compute":** to compute the Variance of Logarithms. If you also want the standard deviation of this index, choose the option for computing with a standard deviation.

Case 2: Two distributions

To compute the Variance of Logarithms of two distributions:

1- From the main menu, choose the item: "Inequality⇒ Variance of Logarithms ".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_1	Optional

Among the buttons, you will find the command « **Compute ».** To compute the standard deviation of this index, choose the option for computing with standard deviation.

The Relative Mean Deviation Index

Denote the Relative Mean Deviation index of inequality for the group k by RMD. It can be expressed as follows:

$$RMD = \frac{\sum_{i=1}^{n} w_i^k \left| (y_i / \mu) - 1 \right|}{\sum_{i=1}^{n} w_i^k}$$

Case 1: One distribution

If you wish to compute the relative mean deviation index of inequality for only one distribution, follow these steps:

1- From the main menu, choose the following items "Inequality⇒ Relative Mean Deviation ".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and values of parameters as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional

Among the buttons, you will find:

- "**Compute":** to compute the relative mean deviation. If you also want the standard deviation of this index, choose the option for computing with a standard deviation.

Case 2: Two distributions

To compute the relative mean deviation of two distributions:

1- From the main menu, choose the item: "Inequality⇒ Relative Mean Deviation ".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_1	Optional

Among the buttons, you will find the command « **Compute ».** To compute the standard deviation of this index, choose the option for computing with standard deviation.

The Conditional Mean Ratio

Denote the Conditional Mean for group k by $\mu(k;p_1;p_2)$, where p_1 and p_2 specify the percentile (p) range of those we wish to include in the computation of the conditional mean. These percentile values p are such that $p_1 \le p \le p_2$. $\mu(k;p_1;p_2)$ is formally defined as:

$$\mu(k;p_1;p_2) = \frac{\int_{p_1}^{p_2} Q(k;p)dp}{p_2 - p_1}$$

and is the average income of those whose rank in the population is between p_1 and p_2. The Conditional Mean Ratio for group k is then given by CMR(k_1,k_2;,p_1,p_2,p_3,p_4) and is defined as

$$CMR(k_1,k_2;p1,p2,p3,p4) = \frac{\mu(k_1;p_1;p_2)}{\mu(k_2;p_3;p_4)}$$

Case 1 : One distribution

If you wish to compute the Conditional Mean Ratio index of inequality for only one distribution, follow these steps:

1- From the main menu, choose "Inequality⇒ Conditional Mean Ratio index".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Percentile	p_1	Compulsory
Percentile	p_2	Compulsory
Percentile	p_3	Compulsory
Percentile	p_4	Compulsory

Among the buttons., you will find the following command:

- "**Compute":** to compute the Conditional Mean Ratio. If you also want the standard deviation of this index, choose the option for computing with a standard deviation.

Case 2: Two distributions

To compute the Conditional Mean Ratio with two distributions:

1- From the main menu, choose the item: "Inequality⇒ Conditional Mean Ratio index".
2- In the configuration of application, choose 2 for the number of distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
percentile	p_1	p_1	Compulsory
percentile	p_2	p_2	Compulsory
percentile	p_3	p_3	Compulsory
percentile	p_4	p_4	Compulsory

Among the buttons, you will find the command « **Compute ».** To compute the standard deviation of this index, choose the option for computing with standard deviation.

The Gini Impact of Component Growth

Let J components y^j add up to y, that is:

$$y_i = \sum_{j=1}^{J} y_i^j$$

The S-Gini index of inequality can be expressed as follows:

$$I(\rho) = \sum_{j=1}^{J} \frac{\mu_j}{\mu_y} IC_j \ (\rho)$$

The contribution of the j^{th} component to total inequality in y is $\frac{\mu_j}{\mu_y} IC_j(\rho)$, where $IC_j(\rho)$ is the coefficient of concentration of the j^{th} component and μ_j is the mean of that component.

The impact on the S-Gini index of growth in y coming exclusively from growth in the j^{th} component is:

$$\frac{\frac{\partial I(\rho)}{\partial y^j}}{\frac{\partial \mu_y}{\partial y^j}/\mu_y} = IC_j(\rho) - I(\rho)$$

When multiplied by 1%, this says for instance by how much (in absolute, not in percentage, terms) the Gini index will change if total income increases by 1% when that growth is entirely due to growth from the j^{th} component. If you wish to compute this statistics, choose from the main menu the following items "Inequality⇒ Impact of Component Growth".

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Component	y^j	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Rho	ρ	Compulsory

Among the buttons, you will find:

- "**Compute":** to compute the impact on the S-Gini index of growth in y coming exclusively from growth in the j^{th} component. If you also want its standard deviation, choose the option for computing with a standard deviation.

The Gini Component Elasticity

The Gini j^{th} -component elasticity is given by:

$$\left(\frac{\frac{\partial I(\rho)}{\partial y^j}}{\frac{\partial \mu_y}{\partial y^j}}\right) \Big/ \left(\frac{I(\rho)}{\mu_y}\right) = \frac{IC_j(\rho)}{I(\rho)} - 1$$

This give the elasticity of the Gini index with respect to total income, when the change in total income is entirely due to growth from the j^{th} component. To compute this elasticity, choose from the main menu the following items "Inequality⇒ Gini Component Elasticity".

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Component	y^J	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
rho	ρ	Compulsory

Among the buttons, you will find:

- "**Compute":** to compute the Gini component elasticity. To obtain the standard deviation of that estimate, choose the option for computing with a standard deviation.

Poverty indices

DAD offers four possibilities for fixing the poverty line:

1- A deterministic poverty line set by the user.
2- A poverty line equal to a proportion l of the mean.
3- A poverty line equal to a proportion m of a quantile Q(p).
4- An estimated poverty line that is asymptotically normally distributed with a standard deviation specified by the user.

For the first possibility, just indicate the value of the deterministic poverty line in front of the indication "**Poverty line".** For the three other possibilities, proceed as follows:

- Click on the button "**Compute line**".
- Choose one of the three following options:

a) Proportion of mean: *the proportion* l *should be indicated.*
b) Proportion of quantile: *indicate the proportion* m *and the quantile* Q(p) *by specifying the desired percentile* p *of the population.*
c) Estimated line: *indicate the estimate of the poverty line* z *and its standard deviation* stdz.

To compute the poverty line in the case of two distributions:

- Click on the button "**Computate line**".
- Choose one of these three following options:

a) Proportion of mean: *indicate the proportions* l_1 *and* l_2 *for the distributions 1 and 2 respectively*.
b) Proportion of quantile: *indicate the proportions* m_1 *and* m_2*, and specify the desired quantiles by indicating the percentiles of population* p_1 *and* p_2.
c) Estimated line: *indicate the estimates of the poverty lines* z_1 *and* z_2 *and their standard deviations* $stdz_1$ *and* $stdz_2$.

The FGT index

The Foster-Greer-Thorbecke poverty index FGT P(k; z; α) for the population subgroup k is as follows:

$$P(k;z;\alpha) = \frac{1}{\sum_{i=1}^{n} w_i^k} \sum_{i=1}^{n} w_i^k (z - y_i)_+^{\alpha}$$

where z is the poverty line and $x_+ = \max(x,0)$. The normalised index is defined by:

$$\overline{P}(k;z;\alpha) = P(k;z;\alpha)/(z)^{\alpha}$$

Case 1: One distribution

To compute the FGT index:

1- From the main menu, choose the item: " Poverty ⇒ FGT index".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	**Variables or parameters**	**Choice is:**
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
Poverty line	z	Compulsory
alpha	α	Compulsory

4- To compute the normalised index, choose that option in the window of inputs.

Among the buttons, you find:

- The command "**Compute":** to compute the FGT index. To compute the standard deviation of this index, choose the option for computing with standard deviation.
- The command "**Graph1":** to draw the value of the index as a function of a range of poverty lines z. To specify the range (for the horizontal axis), choose the item " Graph Management ⇒ Change range of x " from the main menu.
- The command "**Graph2":** to draw the value of $(FGT)^{1/\alpha}$ as a function of a range of parameter α. To specify such a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.

Case 2: Two distributions

To compute the FGT index with two distributions:

1- From the main menu, choose the item: " Poverty ⇒ FGT index".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
Poverty lines	z_1	z_2	Compulsory
alpha	α_1	α_2	Compulsory

To compute the standard deviation of this index, choose the option for computing with standard deviation.

To compute the normalised index, choose this option in the window of inputs.

The Watts poverty index

The Watts poverty index $PW(k;z)$ for the population subgroup k is defined as:

$$PW(k;z) = -\frac{\sum_{i=1}^{n} w_i^k \left(\log(y_i/z)\right)_+}{\sum_{i=1}^{n} w_i^k}$$

where z is the poverty line and $x_+ = \max(x,0)$.

Case 1: One distribution

To compute the Watts index:

1- From the main menu, choose the item: " Poverty ⇒ Watts index".
2- In the configuration of application, choose 1 for the number of distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
Poverty line	z	Compulsory

Commands:

- The command "**Compute":** to compute the Watts index. To compute the standard deviation, choose the option for computing with standard deviation.
- The command "**Graph":** to draw the value of index according to a range of poverty lines z. To specify such a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.

Case 2: Two distributions

To compute the Watts index with two distributions:

1- From the main menu, choose the item: " Poverty ⇒ Watts index".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is:
	Distribution 1	**Distribution 2**	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
Poverty lines	z_1	z_2	Compulsory

To compute the standard deviation, choose the option for computing with standard deviation.

The S-Gini poverty index

The S-Gini poverty index $P(k;z;\rho)$ for the population subgroup k is defined as:

$$P(k;z;\rho) = z - \sum_{i=1}^{n} \left[\frac{(V_i)^\rho - (V_{i+1})^\rho}{[V_1]^\rho} \right] (z - y_i)_+ \quad \text{and} \quad V_i = \sum_{h=i}^{n} w_h^k$$

where z is the poverty line and $x_+ = \max(x,0)$.

Case 1: One distribution

To compute the S-Gini index:

1- From the main menu, choose the item: " Poverty ⇒ S-Gini index".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
Poverty line	z	Compulsory
rho	ρ	Compulsory

4- To compute the normalised index, choose this option in the window of inputs.

Commands:

- The command "**Compute":** to compute the S-Gini index. To compute the standard deviation, choose the option for computing with standard deviation.
- The command "**Graph":** to draw the value of the index according to a range of poverty lines z. To specify such a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.

Case 2: Two distributions

To compute the S-Gini index with two distributions:

1- From the main menu, choose the item: " Poverty ⇒ S-Gini index".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
Poverty lines	z_1	z_2	Compulsory
rho	ρ_1	ρ_2	Compulsory

The first execution bar contains the command « Compute ». To compute the standard deviation, choose the option for computing with standard deviation.

4- To compute the normalised index, choose this option in the window of inputs.

The Clark, Hemming and Ulph (CHU) poverty index

The poverty index $P(k;z;\varepsilon)$ for the population subgroup k is defined as:

$$P(k;z,\varepsilon)=\begin{cases} z-\left(\dfrac{\sum_{i=1}^{n} w_i^k (y_i^*)^{1-\varepsilon}}{\sum_{i=1}^{n} w_i^k}\right)^{1/(1-\varepsilon)} & \text{if} \quad \varepsilon \neq 1 \quad \text{and} \quad \varepsilon \geq 0 \\ z-\exp\left(\dfrac{\sum_{i=1}^{n} w_i^k \ln y_i^*}{\sum_{i=1}^{n} w_i^k}\right) & \text{if} \quad \varepsilon = 1 \end{cases}$$

where z is the poverty line and $y_i^* = \begin{cases} y_i & \text{if} \quad y_i \leq z \\ z & \text{otherwise} \end{cases}$

Case 1: One distribution

To compute the CHU index:

1- From the main menu, choose the item: "Poverty ⇒ CHU index".
2- In the configuration of application, choose 1 for the number of distributions.

3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
Poverty line	z	Compulsory
epsilon	ε	Compulsory

4- To compute the normalised index, choose this option in the window of inputs.

Commands:

- The command "**Compute":** to compute the CHU index. To compute the standard deviation, choose the option for computing with standard deviation.
- The command "**Graph":** to draw the value of the index according to a range of poverty lines z. To specify such a range for the horizontal axis, choose the item " Graph Management $\Rightarrow$ Change range of x " from the main menu.

Case 2: Two distributions

To compute the CHU index with two distributions:

1- From the main menu, choose the item: " Poverty $\Rightarrow$ CHU index".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
Poverty lines	z_1	z_2	Compulsory
epsilon	ε_1	ε_2	Compulsory

The first execution bar contains the command « Compute ». To compute the standard deviation, choose the option for computing with standard deviation.

The Sen Index

The Sen index of poverty $PS(k;z,\rho)$ for the population subgroup k is defined as:

$$PS = H\left[I + (1-I)G^*\right]$$

$$H = \frac{\sum_{i=i}^{n} w_i^k * I(y_i^k \le z)}{\sum_{i=i}^{n} w_i^k}$$

$$q = \frac{\sum_{i=i}^{n} w_i^k * I(z - y_i^k)_+}{\sum_{i=i}^{n} w_i^k}$$

G^* is the Gini index of inequality among the poor,

and where z is the poverty line and $x_+ = \max(x,0)$.

Case 1: One distribution

To compute the Sen index:

1- From the main menu, choose the item: "Poverty ⇒ Sen index".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	*y*	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
Poverty line	z	Compulsory
rho	ρ	Compulsory

4- To compute the normalised index, choose this option in the window of inputs.

Commands:

- The command "**Compute":** to compute the Sen index. To compute the standard deviation, choose the option for computing with standard deviation.
- The command "**Graph":** to draw the value of the index according to a range of poverty lines z. To specify such a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.

Case 2: Two distributions

To compute the Sen index with two distributions:

1- From the main menu, choose the item: "Poverty ⇒ Sen index".
2- In the configuration of application, choose 2 for the number of distributions.
3- Choose the different vectors and parameter values as follows:

Indication	**Vectors or parameters**		**Choice is:**
	Distribution 1	**Distribution 2**	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
Poverty lines	z_1	z_2	Compulsory
rho	ρ_1	ρ_2	Compulsory

4- To compute the normalised index, choose this option in the window of inputs.

The Bi-dimensional FGT index

The Foster-Greer-Thorbecke poverty index for a good g, $P_g(k; z; \alpha)$, for the population subgroup k is as follows:

$$P_g(k;z^g;\alpha) = \frac{\sum_{i=1}^{n} w_i^k (z^g - x_i^g)_+^\alpha}{\sum_{i=1}^{n} w_i^k}$$

where z^g is the poverty line for good g, and $t_+ = \max(t,0)$. The normalised index is defined by:

$$\overline{P}_g(k;z^g;\alpha) = P_g(k;z^g;\alpha)/(z^g)^\alpha$$

- Union headcount

The union headcount, based on G dimensions or commodities, is equal to:

$$P(k;z^1,z^2,...) = \frac{\sum_{i=1}^{n} w_i^k \left(1 - \prod_{g=1}^{G} I(z^g < x_i^g)\right)}{\sum_{i=1}^{n} w_i^k}$$

- Intersection headcount

The intersection headcount, based on G dimensions or commodities, is equal to:

$$P(k;z^1,z^2,...) = \frac{\sum_{i=1}^{n} w_i^k \prod_{g=1}^{G} I(z^g \geq x_i^g)}{\sum_{i=1}^{n} w_i^k}$$

- Union sum of gaps

The union sum of gaps, using G dimensions or commodities, is equal to:

$$P(k;z^1,z^2,....) = \frac{\sum_{i=1}^{n} w_i^k \left(\sum_{g=1}^{G} (z^g - x_i^g)_+\right)}{\sum_{i=1}^{n} w_i^k}$$

- Intersection sum of gaps

The intersection sum of gaps, using G dimensions or commodities, is equal to:

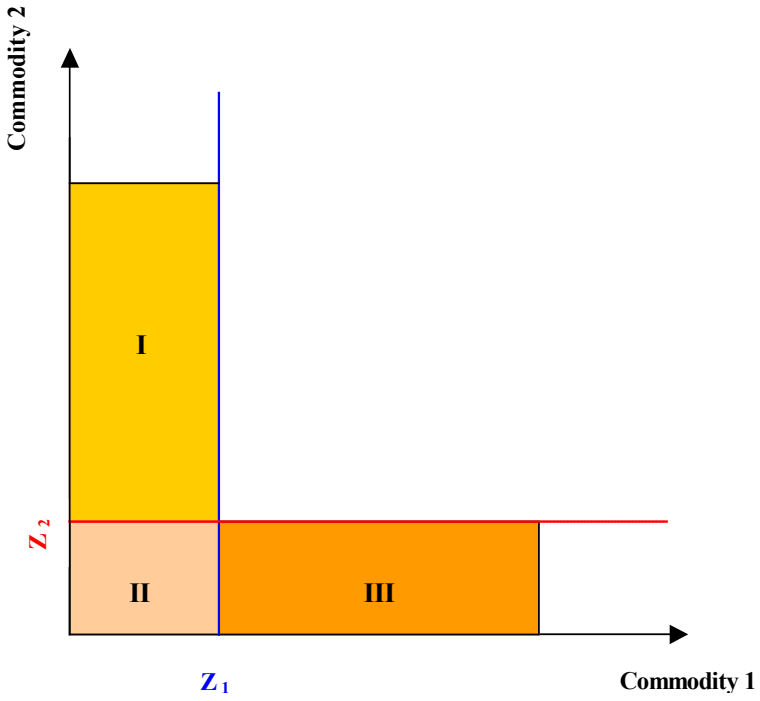
$$P(k;z^1,z^2,...) = \frac{\sum_{i=1}^{n} w_i^k \left(\sum_{g=1}^{G} (z^g - x_i^g)_+ * \prod_{i=1}^{G} I(z^g \geq x_i^g)\right)}{\sum_{i=1}^{n} w_i^k}$$

- Intersection product of gaps

The intersection product of gaps, using G dimensions or commodities, is equal to:

$$P(k;z^1,z^2,...;\alpha_1,\alpha_2,...)=\frac{\sum_{i=1}^{n} w_i^k\left(\prod_{g=1}^{G}(z^g-x_i^g)^{\alpha_g}{}_+ * \prod_{i=1}^{G} I(z^g \geq x_i^g)\right)}{\sum_{i=1}^{n} w_i^k}$$

Graphical illustration for two commodities



Case 1: One distribution

To compute the bi-dimensional FGT indices for two goods:

1- From the main menu, choose the item: " Poverty ⇒ Bidimensional FGT index".
2- Choose the different vectors and parameter values as follows:

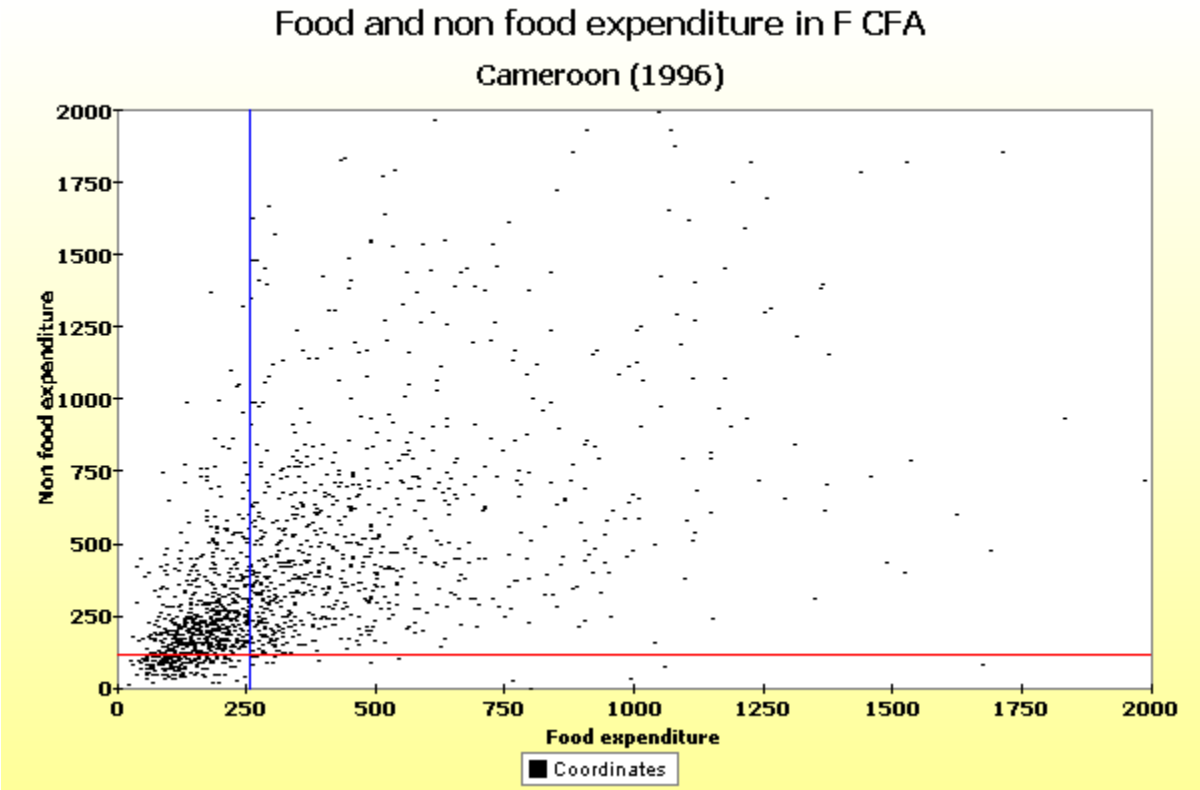
Indication	Variables or parameters	Choice is:
Commodity	x^1	Compulsory
Commodity	x^2	Compulsory

Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Poverty line 1	z^1	Compulsory
Poverty line 2	z^2	Compulsory
alpha1	α_1	Compulsory
alpha2	α_2	Compulsory

Results of this application are:

- *FGT index for commodity 1*: corresponding to areas (I+II) in the graphical illustration.
- *FGT index for commodity 2*: corresponding to areas (II+III) in the graphical illustration.
- *FGT index for the two commodities* (Union approach): corresponding to areas (I+II+III) in the graphical illustration.
- *FGT index for the two commodities* (Intersection approach): corresponding to areas (II) in the graphical illustration.

Example: Food and non-food expenditures per day in F CFA (Cameroon 1996). Food poverty line evaluated at 256 F CFA and non-food poverty line evaluated at 117 F CFA.



Case 2: Two distributions

To compute the FGT indices for two goods and for two distribution:

1- From the main menu, choose the item: " Poverty ⇒ Two Dimensions FGT index ".
2- In the configuration of application, choose 2 for the number of distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Commodity	x_1	x_1	Compulsory
Commodity	x_2	x_2	Compulsory
Size variable	s1	S2	Optional
Group Variable	c	c	Optional
Group Number	k	k	Optional
Poverty line 1	z^1	Z^1	Compulsory
Poverty line 2	z^2	z^2	Compulsory
alpha1	α_1	α_1	Compulsory
alpha2	α_2	α_2	

Impact of a price change on the FGT index

The impact of a good 1's marginal price change (denoted IMP) on the FGT poverty index P(k; z; α) is as follows:

$$IMP = \frac{\partial P(k;z;\alpha)}{\partial p_1} * pc$$

$$= CD^{\alpha+1}{}_1(k;z) * pc$$

where z is the poverty line, k is the population subgroup for which we wish to assess the impact of the price change, and pc is the percentage price change for good l.

$$IMP = \begin{cases} \dfrac{\alpha}{\sum_{i=1}^{n} w_i^k z^\alpha} \sum_{i=1}^{n} w_i^k \left(\dfrac{z - y_i}{z} \right)_+^{\alpha-1} x_i^l & \text{if} \quad \alpha \geq 1 \quad \text{and Normalised} \\ \dfrac{\alpha}{\sum_{i=1}^{n} w_i^k} \sum_{i=1}^{n} w_i^k (z - y_i)_+^{\alpha-1} x_i^l & \text{if} \quad \alpha \geq 1 \quad \text{and Not Normalised} \\ E\left[x^l \mid y = z\right] * f(z) = \dfrac{\sum_{i=1}^{n} w_i^k K_h(z - y_i) * x_i^l}{\sum_{i=1}^{n} w_i^k} & \text{if} \quad \alpha = 0 \end{cases}$$

where x_i^l is expenditure on commodity l by individual i, and $f_+ = \max(f, 0)$. Note that if the FGT index is normalized: $IMP = CD^{\alpha+1}{}_l(k; z) * pc$

To compute the impact of the price change:

1- From the main menu, choose the item: " Poverty ⇒ Impact of price change".
2- Choose the different vectors and parameter values as follows:

Indication	**Variables or parameters**	**Choice is:**
Variable of interest	y	Compulsory
Size variable	s	Optional
Commodity	x	Compulsory
Group Variable	c	Optional
Group Number	k	Optional
Poverty line	z	Compulsory
alpha	α	Compulsory
Price change in %	pc	Compulsory

Commands:

- "**Compute":** to compute the impact of the price change. To compute the standard deviation of this estimated impact, choose the option for computing with standard deviation.
- "**Graph":** to draw the value of the impact as a function of a range of poverty lines z. To specify that range (and thus the range of the horizontal axis), choose the command "**Range"**.

Impact of a tax reform on the FGT indices

This tax reform consists of a variation in the prices of two commodities 1 and 2, under the constraint that it leaves unchanged total government revenue. The effect of this constraint is given by an efficiency parameter, "gamma" (γ), which is the ratio of the marginal cost of public funds (MCPF) from a tax on 2 over the MCPF from a tax on 1.

The impact of this tax reform (denoted IMTR) on the FGT poverty index P(k; z; α) is as follows:

$$IMTR = \left[CD_1^{\alpha+1}(k;z) - \gamma \frac{\overline{X}_1}{\overline{X}_2} CD_2^{\alpha+1}(k;z) \right] * pc$$

where z is the poverty line, $CD_1^{\alpha+1}(k;z)$ and $CD_2^{\alpha+1}(k;z)$ are the consumption dominance curves of commodities 1 and 2, and pc is the percentage price change of commodity 1. Under the government revenue constraint, the percentage price change of commodity 1 is given by $\gamma \frac{\overline{X}_1}{\overline{X}_2} pc$.

To compute the impact of the tax reform:

1- From the main menu, choose the item: " Poverty $\Rightarrow$ Impact of tax reform".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Commodity 1	x_1	Compulsory
Commodity 2	x_2	Compulsory
Group Variable	c	Optional
Group Number	k	Optional
Poverty line	z	Compulsory
alpha	α	Compulsory

gamma	γ	Compulsory
1' s % price change	pc	Compulsory

Commands:

- "**Compute":** to compute the impact of the tax reform . To compute the standard deviation of this estimated impact, choose the option for computing with standard deviation.
- "**Critical γ ":** to compute the gamma at which the tax reform will have zero impact on poverty. The value of this critical gamma equals $\overline{CD}_1^{\alpha+1}(k;z)/\overline{CD}_2^{\alpha+1}(k;z)$
- "**Graph z":** to draw the value of the impact of the tax reform as a function of a range of poverty lines z. To specify that range (and the horizontal axis), choose the command "**Range"**.
- " **Graph γ ":** to draw the value of the impact as a function of a range of MCPF ratios γ . To specify that range (and the horizontal axis), choose the command "**Range"**.

Lump-sum Targeting

The per-capita-dollar impact of a marginal addition of a constant amount of income to everyone within a group k – called Lump-Sum Targeting (LST) – on the FGT poverty index P(k; z; α), is as follows:

$$LST = \begin{cases} -\alpha P(k,z;\alpha-1) & \text{if} \quad \alpha \geq 1 \quad \text{and} \quad \text{Not Normalised} \\ -\dfrac{\alpha}{z}\overline{P}(k,z;\alpha-1) & \text{if} \quad \alpha \geq 1 \quad \text{and} \quad \text{Normalised} \\ -f(k,z) & \text{if} \quad \alpha = 0 \end{cases}$$

where z is the poverty line, k is the population subgroup for which we wish to assess the impact of the income change, and f(k,z) is the density function of the group k at level of income z.

To compute that impact:

1- From the main menu, choose the item: " Poverty $\Rightarrow$ Lump-sum Targeting".
2- Choose the different vectors and parameter values as follows:

Indication	**Variables or parameters**	**Choice is:**
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional

Group Number	k	Optional
Poverty line	z	Compulsory
alpha	α	Compulsory

Commands:

- "**Compute":** to compute the impact of the income change. To compute the standard deviation of this estimated impact, choose the option for computing with standard deviation.
- "**Graph":** to draw the value of the impact as a function of a range of poverty lines z. To specify that range (and thus the range of the horizontal axis), choose the command "**Range"**.

Inequality-neutral Targeting

The per-capita-dollar impact of a proportional marginal variation of income for the group k, called Inequality Neutral Targeting, on the FGT poverty index P(k; z; α) is as follows:

$$INT = \begin{cases} \alpha \dfrac{P(k,z;\alpha) - zP(k,z;\alpha-1)}{\mu_k} & \text{if} \quad \alpha \geq 1 \quad \text{and} \quad \text{FGT is not normalised} \\ \alpha \dfrac{\overline{P}(k,z;\alpha) - z\overline{P}(k,z;\alpha-1)}{\mu_k} & \text{if} \quad \alpha \geq 1 \quad \text{and} \quad \text{FGT is normalised} \\ -\dfrac{zf(k,z)}{\mu_k} & \text{if} \quad \alpha = 0 \end{cases}$$

where z is the poverty line, k is the population subgroup for which we wish to assess the impact of the income change, and f(k,z) is the density function of the group k at level of income z.

To compute that impact:

1- From the main menu, choose the item: " Poverty ⇒ Inequality-neutral Targeting".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	w	Optional
Group Variable	c	Optional
Group Number	k	Optional
Poverty line	z	Compulsory
alpha	α	Compulsory

Commands:

- "**Compute":** to compute the impact. To compute the standard deviation of this estimated impact, choose the option for computing with standard deviation.
- "**Graph":** to draw the value of the impact as a function of a range of poverty lines z. To specify that range (and thus the range of the horizontal axis), choose the command "**Range"**.

Growth Elasticity

The overall growth elasticity (GREL) of poverty, when growth comes exclusively from growth within a group k (which is, within that group, inequality neutral), is given by:

$$\text{GREL} = \begin{cases} \alpha \dfrac{P(k,z;\alpha) - zP(k,z;\alpha-1)}{P(z,\alpha)} & \text{if} \quad \alpha \geq 1 \\ -\dfrac{zf(k,z)}{F(z)} & \text{if} \quad \alpha = 0 \end{cases}$$

where z is the poverty line, k is the population subgroup in which growth takes place, f(z) is the density function at level of income z, and F(z) is the headcount.

To compute that growth elasticity:

1- From the main menu, choose the item: " Poverty $\Rightarrow$ Growth Elasticity".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory

Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Poverty line	z	Compulsory
alpha	α	Compulsory

Commands:

- "**Compute":** to compute the growth elasticity. To compute the standard deviation of its estimate, choose the option for computing with standard deviation.
- "**Graph":** to draw the value of the impact as a function of a range of poverty lines z. To specify that range (and thus the range of the horizontal axis), choose the command "**Range"**.

The Impact of Component Growth

The per-capita-dollar impact of growth in the j^{th} component on the normalized FGT index of the k^{th} group is as follows:

$$\frac{\frac{\partial \overline{P}(k;z,\alpha)}{\partial y^j}}{\frac{\partial \mu}{\partial y^j}} = -\overline{CD}^j(k;z,\alpha)$$

where $\overline{CD}$ is the normalized C-dominance curve of the component j. If you wish to compute that impact, choose "Poverty⇒ Impact of Component Growth".

Indication	**Variables or parameters**	**Choice is:**
Variable of interest	y	Compulsory
Income Component	y^j	Compulsory
Size variable	w	Optional
Group Variable	c	Optional
Group Number	k	Optional
Alpha	α	Compulsory
Poverty line	z	Compulsory

Among the buttons, you will find:

- "**Compute":** to compute the statistics. If you also want its standard error, choose the option for computing with a standard deviation.

The Component Elasticity of Poverty

The j^{th} component elasticity of poverty (measured by the normalized FGT index) is:

$$-\frac{\mu}{\overline{P}(k;z,\alpha)}\overline{CD}^{j}(k;z,\alpha)$$

where $\overline{CD}^{j}$ is the normalized C-dominance curve of the component j. If you wish to compute this elasticity choose "Poverty⇒ Component Elasticity".

Indication	**Variables or parameters**	**Choice is:**
Variable of interest	y	Compulsory
Income Component	y^j	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Alpha	α	Compulsory
Poverty line	z	Compulsory

Among the buttons, you will find:

- "**Compute":** to compute the statistics. To obtain the standard deviation, choose the option for computing with a standard deviation.

The social welfare indices

DAD can compute the following types of social welfare indices:

The Atkinson social welfare index

Case 1: One distribution

To compute the Atkinson index of social welfare for one distribution:

1- From the main menu, choose the following item: "Welfare ⇒ Atkinson index".
2- In the configuration of the application, choose 1 for the number of distributions.
3- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
epsilon	ε	Compulsory

Commands:

- The command "**Compute**": to compute the Atkinson index. To compute the standard deviation, choose the option for computing with standard deviation.
- The command "**Graph**": to draw the value of the index according to a range of parameters ε. To specify such a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.

Case 2: Two distributions

To compute the Atkinson with two distributions:

1- From the main menu, choose the item: "Welfare ⇒ Atkinson index".
2- In the configuration of application, choose 2 for the number of distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
epsilon	ε_1	ε_2	Compulsory

To compute the standard deviation, choose the option for computing with standard deviation.

The S-Gini social welfare index

Case1: One distribution

To compute the S-Gini index of social welfare for one distribution:

1- From the main menu, choose the following item: "Welfare ⇒ S-Gini index".
2- In the configuration of the application, choose 1 for the number of distributions.
3- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
rho	ρ	Compulsory

Commands:

- The command **"Compute"**: to compute the S-Gini index. To compute the standard deviation, choose the option for computing with standard deviation.
- The command **"Graph"**: to draw the value of the index according to a range of parameter ρ. To specify such a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.

Case 2 : Two distribution

To compute the S-Gini with two distributions:

1- From the main menu, choose the item: "Welfare ⇒ S-Gini index".
2- In the configuration of application, choose 2 for the number of distributions.

3- Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
rho	ρ_1	ρ_2	Compulsory

To compute the standard deviation, choose the option for computing with standard deviation.

The Atkinson-Gini social welfare index

To compute the Atkinson-Gini social welfare index:

1- From the main menu, choose the following item: "Welfare ⇒ S-Gini index".
2- In the configuration of the application, choose 1 for the number of distributions.
3- After confirming the configuration, the application appears. Choose the different vectors and values of parameters as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
epsilon	ε	Compulsory
rho	ρ	Compulsory

Press the command "**Compute**" to compute the Atkinson-Gini index. To compute the standard deviation, choose the option for computing with standard deviation.

Case 2: Two distributions

To compute the Atkinson-Gin social welfare with two distributions:

1- From the main menu, choose the item: "Welfare ⇒ Atkinson-Gini".
2- In the configuration of application, choose 2 for the number of distributions.

3- Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
rho	ρ_1	ρ_2	Compulsory
epsilon	ε_1	ε_2	Compulsory

To compute the standard deviation, choose the option for computing with standard deviation.

Impact of a price change on the Atkinson Social Welfare Index

The impact of a good 1's marginal price change (denoted IMPW) on the Atkinson Social Welfare index $\xi(\varepsilon)$ is as follows:

$$\text{IMPW} = \frac{\partial\, \xi(\varepsilon)}{\partial\, p_1} * \text{pc}$$

$$\text{IMPW} = \begin{cases} -(s1)^{\frac{1}{\varepsilon-1}} * (s2)^{\frac{\varepsilon}{1-\varepsilon}} * (s3) * \text{pc} & \text{if} \quad \varepsilon \neq 1 \\ -\exp(s2/s1) * s3/s1 * \text{pc} & \text{if} \quad \varepsilon \neq 1 \end{cases}$$

and

$$\begin{cases} s1 = \sum_i w_i \quad s2 = \sum_i w_i y_i^{1-\varepsilon} \quad s3 = \sum_i w_i y_i^{-\varepsilon} x_i & \text{if} \quad \varepsilon \neq 1 \\ s1 = \sum_i w_i \quad s2 = \sum_i w_i \log(y_i) \quad s3 = \sum_i w_i x_i / y_i & \text{if} \quad \varepsilon = 1 \end{cases}$$

where x_i^1 is expenditure on commodity 1 by individual i, y_i is the variable of interest ("living standard"), and pc is the percentage price change for good 1.

To compute the impact of the price change:

1- From the main menu, choose: " Welfare ⇒ Impact of price change".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Commodity	x	Compulsory
Group Variable	c	Optional
Group Number	k	Optional
epsilon	ε	Compulsory
Price change in %	pc	Compulsory

The computation can be made solely within a group of individuals. This is done by specifying the group number k and the group variable c.

Commands:

- **"Compute":** to compute the impact of the price change. To compute the standard deviation of this estimated impact, choose the option for computing with standard deviation.
- **"Graph":** to draw the value of the impact as a function of a range for the parameter ε. To specify that range (and thus the range of the horizontal axis), choose the command "**Range**".

Impact of a tax reform on the Atkinson Social Welfare Index

This tax reform consists of a variation in the prices of two commodities 1 and 2, under the constraint that it leaves unchanged total government revenue. The effect of this constraint is given by an efficiency parameter, "gamma" (γ), which is the ratio of the marginal cost of public funds (MCPF) from a tax on 2 over the MCPF from a tax on 1.

The impact of this tax reform (denoted IMWTR) on the Atkinson Social Welfare index $\xi(\varepsilon)$ is as follows:

$$\text{IMWTR} = \left[\frac{\partial\, \xi(\varepsilon)}{\partial\, p_1} - \gamma \frac{\overline{X}_1}{\overline{X}_2} \frac{\partial\, \xi(\varepsilon)}{\partial\, p_2} \right] * \text{pc}$$

where pc is the percentage price change of commodity 1, and $\overline{X}_g$ is the total expenditure on the good g. Under the government revenue constraint, the percentage price change of commodity 1 is given by $\gamma \frac{\overline{X}_1}{\overline{X}_2} \text{pc}$. The computation can be made solely within a group of individuals. This is done by specifying the group number k and the group variable c.

To compute the impact of the tax reform:

1- From the main menu, choose " Welfare ⇒ Impact of tax reform".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Commodity 1	x_1	Compulsory
Commodity 2	x_2	Compulsory
Group Variable	c	Optional
Group Number	k	Optional
epsilon	ε	Compulsory
gamma	γ	Compulsory
1' s % price change	pc	Compulsory

Commands:

- **"Compute":** to compute the impact of the tax reform . To compute the standard deviation of this estimated impact, choose the option for computing with standard deviation.

Impact of Income-component growth on the Atkinson Social Welfare Index

The impact of growth in the j^{th} component on the Atkinson Social Welfare index $\xi(\varepsilon)$ is as follows:

$$\frac{\partial \xi(\varepsilon)}{\partial x_j} * pc = \begin{cases} (s1)^{\frac{1}{\varepsilon-1}} * (s2)^{\frac{\varepsilon}{1-\varepsilon}} * (s3) * pc & \text{if} \quad \varepsilon \neq 1 \\ \exp(s2/s1) * s3/s1 * pc & \text{if} \quad \varepsilon \neq 1 \end{cases}$$

and

$$\begin{cases} s1 = \sum_i w_i \quad s2 = \sum_i w_i y_i^{1-\varepsilon} \quad s3 = \sum_i w_i y_i^{-\varepsilon} x_i^j & \text{if} \quad \varepsilon \neq 1 \\ s1 = \sum_i w_i \quad s2 = \sum_i w_i \log(y_i) \quad s3 = \sum_i w_i x_i / y_i & \text{if} \quad \varepsilon = 1 \end{cases}$$

where x_i^j is the value of component j for individual i and pc is the percentage change in that j income component. This tells us therefore by how much social welfare will change if a growth of pc is observed in a component j of total income.

To compute the impact of that change:

1- From the main menu, choose the item: " Welfare ⇒ Impact of Income-component growth".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Component	x	Compulsory
Group Variable	c	Optional
Group Number	k	Optional
Epsilon	ε	Compulsory
Component change in %	pc	Compulsory

Commands:

- "**Compute":** to compute the impact of the Income-component growth. To compute the standard deviation of this estimated impact, choose the option for computing with standard deviation.
- "**Graph":** to draw the value of the impact as a function of a range for parameter ε. To specify that range (and thus the range of the horizontal axis), choose the command "**Range"**.

The decomposition of inequality and poverty

The decomposition of the FGT index

The FGT poverty index for a population composed of K groups can be written as follows:

$$P(z;\alpha) = \sum_{k=1}^{K} \phi(k)P(k;z;\alpha)$$

where $P(k;z;\alpha)$ is the FGT poverty index for subgroup k and $\phi(k)$ is the proportion of the population in this subgroup. The contribution of group k to the poverty index for the whole population equals $\phi(k)P(k;z;\alpha)$.

To perform the decomposition of the FGT index:

1- From the main menu, choose the item: " Decomposition ⇒ FGT Decomposition".
2- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size Variable	s	Optional
Group Variable	c	Optional
Poverty line	z	Compulsory
alpha	α	Compulsory
Group numbers separated by "-"	k_1 - k_2 -…	Compulsory

Remark:

*The group numbers separated by the dash "-" should be integer values. For example, we may have two subgroups coded by the integers 1 and 2. In this case, we would write in the field « **Group Numbers** » the values "1-2" before proceeding to the decomposition.*

The decomposition of the FGT index for two groups

To perform the decomposition of the FGT index for two groups:

1- From main menu, choose the item: "Decomposition ⇒ FGT Decomposition for two groups".
2- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size Variable	s	Optional
Group Variable	c	Optional
Poverty line	z	Compulsory
alpha	α	Compulsory
Numbers for the 2 subgroups separated by "-"	k_1 - k_2	Compulsory

In the output window, you will find the following information:

1- The FGT index for the whole population.
2- The FGT index for each of the two subgroups.
3- The difference in the indices of the two groups: $P(1;z;\alpha) - P(2;z;\alpha)$
4- The percentage difference in the contribution of the two population subgroups, $(\phi(1)P(1;z;\alpha) - \phi(2)P(2;z;\alpha)) / P(z;\alpha)$

To compute the standard deviations for these statistics, choose the option computing with standard deviation.

The decomposition of the FGT index across growth and redistribution effects

We can decompose variation of the FGT Index between two periods, t1 and t2, into growth and redistribution effects as follows:

$$\underbrace{P_2 - P_1}_{\text{Variation}} = \underbrace{\left[P(\mu^{t2}, \pi^{t1}) - P(\mu^{t1}, \pi^{t1})\right]}_{C1} + \underbrace{\left[P(\mu^{t1}, \pi^{t2}) - P(\mu^{t1}, \pi^{t1})\right]}_{C2} + R$$

Variation = Difference in poverty between t1 and t2.
C1 = Growth Impact.
C2 = Contribution of redistribution effect
R = Residual

$P(\mu^{t2}, \pi^{t1})$: the FGT index of the first period when we multiply all incomes y_i^{t1} of the first period by the ratio μ^{t2} / μ^{t1}

$P(\mu^{t1}, \pi^{t2})$. the FGT index of the second period when we multiply all incomes y_i^{t2} of the second period by the ratio μ^{t1} / μ^{t2}

To perform the decomposition of the FGT index across growth and redistribution effects:

1- From the main menu, choose the item: "Decomposition ⇒ Growth and redistribution".
2- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	**Vector or parameter**		**Choice is:**
	Distribution - t1	**Distribution-t2**	
Variable of interest	y^1	y^2	Compulsory
Size Variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Index of group	k_1	k_2	Optional
Poverty lines	z		Compulsory
alpha	α		Compulsory

To compute the standard deviation of this index, choose the option for computing with standard deviation.

The sectoral decomposition of differences in FGT indices

We can decompose differences in FGT into sub-group differences in poverty and population proportions as follows:

$$\underbrace{P_2 - P_1}_{\text{Variation}} = \left[\sum_{k=1}^{K} \phi_1(k)\left(P_2(k;z;\alpha) - P_1(k;z;\alpha)\right)\right] + \left[\sum_{k=1}^{K} P_1(k;z;\alpha)\left(\phi_2(k) - \phi_1(k)\right)\right] + \left[\sum_{k=1}^{K} \left(P_2(k;z;\alpha) - P_1(k;z;\alpha)\right)\left(\phi\right.\right.$$

Variation = Difference in poverty between 1 and 2.
C1 = Intra-sectoral or intra-group impacts
C2 = Impact of changes in subgroup proportions
C3 = Interaction effect

To perform this decomposition:

1- From the main menu, choose: "Decomposition ⇒ Sectoral".
2- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size Variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Poverty lines	z		Compulsory
alpha	α		Compulsory
Group numbers separated by "-"	k_1-k_2-…		Compulsory

To compute the standard deviation of this index, choose the option for computing with standard deviation.

The impact of demographic changes

This application computes the impact of a change (by a given percentage) in the proportion of a group t. That change is accompanied by an exactly offsetting change in the proportion of the other groups.

If the population proportion of group t increases by pc percent, such that $\phi(t) \rightarrow (\phi(t)(1+pc))$, the total estimated impact on poverty is as follows:

$$\Delta P = \left(\phi(t) * P(t;z,\alpha) - \sum_{k \neq s}^{K} \frac{\phi(t)}{1-\phi(t)} * \phi(k) * P(k;z,\alpha) \right) * pc$$

If the population proportion of group s increases by absolute pc percent of the total population, such that $\phi(t) \rightarrow (\phi(t)+pc)$, the total estimated impact on poverty is as follows:

$$\Delta P = \left(P(t;z,\alpha) - \sum_{k \neq s}^{K} \frac{\phi(k)}{1-\phi(t)} * P(k;z,\alpha) \right) * pc$$

where $P(k;z;\alpha)$ is the FGT poverty index for subgroup k and $\phi(k)$ is the proportion of the population found in that subgroup.
To perform this estimation:

1- From the main menu, choose: " Decomposition ⇒ Impact of Demographic Change".
2- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size Variable	s	Optional
Group Variable	c	Optional
Changed group	t	Compulsory
Poverty line	z	Compulsory
Alpha	α	Compulsory
Group numbers separated by "-"	k_1-k_2-…	Compulsory

Remark:

*The group numbers separated by the dash "-" should be integer values. For example, we may have two subgroups coded by the integers 1 and 2. In this case, we would write in the field « **Group Numbers** » the values "1-2" before proceeding to the decomposition.*

The decomposition of the S-Gini index of inequality

Let J components y^j add up to y, that is:

$$y_i = \sum_{j=1}^{J} y_i^j$$

We can decompose the S-Gini index of inequality as follows:

$$I(\rho) = \sum_{j=1}^{J} \frac{\mu_j}{\mu} IC_j\ (\rho)$$

The contribution of the j^{th} component to inequality in y is $\frac{\mu_j}{\mu_y} IC_j(\rho)$,

where $IC_j(\rho)$ is the coefficient of concentration of the j^{th} component and μ_j is the mean of that component.

To perform the decomposition of the S-Gini index of inequality:

1- From the main menu, choose the item: "Welfare and inequality ⇒ Decomposition ⇒ S-Gini decomposition".
2- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Size Variable	s	Optional
rho	ρ	Compulsory
Vector(s) of interest	Index1-index2…	Compulsory

The following results appear in the output window:

1- The S-Gini index for y.
2- The coefficients of concentration for every component of y.
3- The ratio μ_j / μ for every component of y.
4- The contribution for every component.

The decomposition of the Generalised Entropy index of inequality

The Generalised Entropy index of inequality can be decomposed as follows:

$$I(\theta) = \sum_{k=1}^{K} \phi(k) \left(\frac{\mu(k)}{\mu_y} \right)^{\theta} .I(k;\theta) + \bar{I}(\theta)$$

where:

$\phi(k)$ is the proportion of the population found in subgroup k.
$\mu(k)$ is the mean income of group k.
$I(k;\theta)$ is the inequality within group k.
$\bar{I}(\theta)$ is population inequality if each individual in subgroup k is given the mean income of subgroup k, $\mu(k)$.

To perform the decomposition of the entropy index:

1- From the main menu, choose the item : "Welfare and inequality ⇒ Decomposition ⇒ Entropy decomposition".
2- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size Variable	s	Optional
Group Variable	c	Optional
theta	θ	Compulsory
Group numbers separated by "-"	k_1-k_2-…	Compulsory

The following information appears in the output window:

1- The entropy index for the whole population.
2- The entropy index for between-group inequality $\bar{I}(\theta)$.
3- The entropy index within every subgroup $I(k;\theta)$.
4- The ratio $(\mu(k)/\mu)$ "Normalised mean" for every subgroup.
5- The absolute contribution to total inequality of inequality within every subgroup, that is, $(\mu(k)/\mu)^{\theta}.\phi(k).I(k;\theta)$
6- The relative contribution to total inequality of inequality within every subgroup.

To compute the standard deviations for these statistics, choose the option computing with standard deviation.

Decomposition of variation of social welfare index between two periods

We can decompose the difference in social welfare (as measured by the EDE Atkinson index) between two populations, 1 and 2, as follows:

$$\xi_2(\varepsilon)-\xi_1(\varepsilon)=\underbrace{(I_1-I_2)*\mu_1}_{C1}+\underbrace{(\mu_2-\mu_1)*(1-I_1)}_{C2}+\underbrace{(\mu_2-\mu_1)*(I_2-I_2)}_{C3}$$

where:
C1: Impact of change in inequality.
C2: Impact of change in mean.
C3: Interaction impact.

To perform this decomposition:

1- From the main menu, choose: "Decomposition ⇒ Decomposition of Social Welfare".
2- Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is:
	Distribution 1	**Distribution 2**	
Variable of interest	y^1	y^2	Compulsory
Size Variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
epsilon	ε_1	ε_2	Compulsory

To compute the standard deviation, choose the option for computing with standard deviation.

Dominance

This section looks at the primal dominance conditions for ordering poverty and inequality across two distributions of living standards. Corresponding dual dominance conditions are considered in the section on **Curves**.

Poverty dominance

Distribution 1 dominates distribution 2 at order s over the conditional range $[z^-, z^+]$ if only if: $P_1(\zeta;\alpha) > P_2(\zeta;\alpha) \quad \forall \quad \zeta \in [z^-, z^+]$ for $\alpha = s-1$.

This involves comparing stochastic dominance curves at order *s* or FGT curves with $\alpha = s-1$. This application checks for the points at which there is a reversal of the dominance conditions. Said differently, it provides the crossing points of the dominance curves, that is, the values of ζ and $P_1(\zeta;\alpha)$ for which $P_1(\zeta;\alpha) = P_2(\zeta;\alpha)$ when

$\text{sign}(P_1(\zeta-\eta;\alpha) - P_2(\zeta-\eta;\alpha)) = \text{sign}(P_2(\zeta+\eta;\alpha) - P_1(\zeta+\eta;\alpha))$

for a small η.

The crossing points of ζ can also be referred to as "critical poverty lines". To check for the crossing points of the dominance curves of two distributions:

1- From main menu, choose the item: "Dominance ⇒ Poverty Dominance".
2- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	**Vector or parameter**		**Choice is:**
	Distribution 1	**Distribution 2**	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
s	s		Compulsory

Commands:

- "**Compute":** to provide the critical poverty lines and the crossing points of the sample dominance curves. When the option "with STD" is specified, the standard deviation on the estimates of the critical poverty lines and on the estimates of the crossing points of the FGT curves are also given.
- "**Range":** to specify the range of poverty lines over which to check for the presence of critical poverty lines. With this command, you can also specify the incremental step of search for these crossing points.
- "**Graph":** to draw the FGT curves for the two distributions.

Inequality dominance

Distribution 1 dominates distribution 2 in inequality at order s over the conditional range of proportions of the mean $\left[l^{-}, l^{+}\right]$ only if $\overline{P}_1(\lambda\mu_1,\alpha) > \overline{P}_2(\lambda\mu_2,\alpha) \quad \forall \quad \lambda \in \left[l^{-}, l^{+}\right]$ where $\alpha = s-1$

These are normalised stochastic dominance curves at order *s* or normalised FGT curves for $\alpha = s-1$. This application checks for the points at which there is a reversal of the above dominance conditions for inequality orderings. Said differently, it provides the crossing points of the FGT curves, that is, the values of λ and $\overline{P}_1(\lambda\mu_1;\alpha)$ for which $\overline{P}_1(\lambda\mu_1;\alpha) = \overline{P}_2(\lambda\mu_2;\alpha)$ when

$\text{sign}(\overline{P}_1((\lambda-\eta)\mu_1;\alpha) - P_2((\lambda-\eta)\mu_2;\alpha)) = \text{sign}(P_2((\lambda+\eta)\mu_2;\alpha) - P_1((\lambda+\eta)\mu_1;\alpha))$
for a small η.

These crossing points at λ can also be referred to as "critical relative poverty lines", when the poverty lines are a proportion of the mean and when the indices are normalised by the poverty line. To check for those crossing points:

1- From main menu, choose the item: "Dominance ⇒ Inequality Dominance".
2- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_1	Optional
s	s		Compulsory

Commands:

- "**Compute":** to provide the critical relative poverty lines and the crossing points of the sample normalised dominance curves. When the option "with STD" is specified, the standard deviation on the estimates of the critical relative poverty lines and on the estimates of the crossing points of the normalised FGT curves are also given.
- "**Range":** to specify the range of λ over which to check the presence of critical values. With this command, you can also specify the incremental step of search for these crossing points.
- "**Graph":** to draw the normalised FGT curves for the two distributions along values of the parameter λ.

Indirect tax dominance

Taxing commodity 2 is better than taxing commodity 1 at order of dominance s over the conditional range $\left[z^-, z^+\right]$ if only if: $\overline{CD}_1^s(k;\zeta) > \gamma\overline{CD}_2^s(k;\zeta) \quad \forall \quad \zeta \in \left[z^-, z^+\right]$.

These are CD curves of order *s*. If this condition holds, then an increase in the price of good 2, with the benefit of a decrease in the price of good 1, will decrease poverty for poverty lines between z^- and z^+ and for poverty indices of order "s". The ratio of the marginal cost of public funds (MCPF) from a tax on 2 over the MCPF from a tax on 1 is also used to determine whether increasing the tax on 2 for the benefit of decreasing the tax on good 1 can be deemed to be "socially efficient".

This application computes differences between $\overline{CD}_1^s(k;\zeta)$ and $\gamma\overline{CD}_2^s(k;\zeta)$. It also checks for the points at which there is a reversal of the dominance conditions. Said differently, it provides the crossing points of the CD curves, that is, the values of ζ and $\overline{CD}^s(k;\zeta)$ for which $\overline{CD}_1^s(k;\zeta) = \gamma\overline{CD}_2^s(k;\zeta)$ when $\text{sign}(\overline{CD}_1^s(k;\zeta-\eta) - \gamma\overline{CD}_2^s(k;\zeta-\eta)) = \text{sign}(\overline{CD}_2^s(k;\zeta+\eta) - \overline{CD}_1^s(k;\zeta+\eta))$ for a small η. The crossing points of ζ can also be referred to as "critical poverty lines".

Critical values of γ are also provided. These are the minimum of $\overline{CD}_1^{\alpha+1}(k;z)/\overline{CD}_2^{\alpha+1}(k;z)$ over an interval $[z^-, z^+]$ of poverty lines z. It gives the maximum ratio of the MCPF (for commodity 2 over that for commodity 1) up to which taxing commodity 2 can be deemed socially efficient.

To use these functions:

1- From the main menu, choose the item: " Dominance $\Rightarrow$ Indirect tax dominance".
2- Choose the different vectors and parameter values as follows:

Indication	**Variables or parameters**	**Choice is:**
Variable of interest	y	Compulsory
Size variable	s	Optional
Commodity 1	x_1	Compulsory
Commodity 2	x_2	Compulsory
Group Variable	c	Optional
Group Number	k	Optional
Poverty line	z	Compulsory
s	s	Compulsory
gamma	γ	Compulsory

Commands:

- " **Critical z ":** to compute the values of the poverty lines at which the CD curves $\overline{CD}_1^s(k;z)$ and $\gamma\overline{CD}_2^s(k;z)$ cross. To specify a range for a search of crossing points, choose the command "**Range"**.
- " **Critical γ ":** to compute the critical gamma for tax dominance. The range $[z^-, z^+]$ is specified under "**Range"**.
- "**Difference":** to compute the difference $\overline{CD}_1^s(k;z) - \gamma\overline{CD}_2^s(k;z)$.
- " **Graph":** to draw the value of $\overline{CD}_1^s(k;z)$ and $\gamma\overline{CD}_2^s(k;z)$ as a function of a range of poverty lines z. To specify that range, choose the command "**Range"**.
- **"Step":** the value of the incremental steps with which the critical z is searched.

Curves

A number of curves are useful to present a general descriptive view of the distribution of living standards. Many of these curves can also serve to check the robustness of distributive orderings in terms of poverty, inequality, social welfare and equity.

Quantiles and normalised quantiles

Remark: The application for computing normalised quantiles is similar in structure to the one for computing quantiles.

The p-quantile at a percentile p of a continuous population is given by:

$Q(p) = F^{-1}(p)$ where $p = F(y)$ is the cumulative distribution function at *y*.

For a discrete distribution, let the n observations of living standards be ordered, such that $y_1 \le y_2 \le \cdots \le y_i \le y_{i+1} \le \cdots \le y_n$. If $p \in [F(y_i), F(y_{i+1})]$, then we define $Q(p) = y_{i+1}$.

The normalised quantile is defined as $\overline{Q}(p) = Q(p)/\mu$.

Case 1: One distribution

To compute the quantiles of one distribution:

1- From the main menu, choose the item: "Curves ⇒ Quantile".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size Variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
p	p	Compulsory

Commands:

- **"Compute":** to compute the quantile at a point p. To compute the standard deviation, choose the option for computing with standard deviation.

- **"Graph":** to draw the value of the curve according to the parameter p. To specify a range for the horizontal axis (for the p values), choose the item "Graph Management ⇒ Change range of x " from the main menu.

Case 2 : Two distributions

To compute the quantiles of two distributions:

1- From the main menu, choose the item: "Curves ⇒ Quantile".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size Variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
p	p_1	p_2	Compulsory

Commands:

- **"Crossing"**: to check if the two quantile curves intersect. If the two curves intersect, DAD indicates the co-ordinates of the first intersection and their standard deviation if the option of computing with standard deviation is chosen. To seek an intersection over a particular range of p, use "**Range**" to specify this range.
- **"Difference" :** to compute the difference $Q_1(p_1) - Q_2(p_2)$.
- **"Graph" :** to draw the difference $Q_1(p) - Q_2(p)$ along values of the parameter p.
- **"Range":** to specify the range for the search for a crossing of the two curves. also specifies the range of the horizontal axis.

Poverty Gap Curve

The poverty gap quantile at a percentile p is:

$$g(p;z) = (z - Q(p))_+$$

Case 1: One distribution

To compute the poverty gap quantile for one distribution:

1- From the main menu, choose the item: "Curves ⇒ Poverty gap quantile".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size Variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Poverty line	z	Compulsory
p	p	Compulsory

Commands:

- "**Compute**": to compute $g(p;z)$. To compute the standard deviation, choose the option for computing with standard deviation.
- "**Graph**": to draw the value of $g(p;z)$ as a function of p. To specify a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.

To compute the standard deviation, choose the option for computing with standard deviation.

Case 2: Two distributions

To reach the application for two distributions:

1- From the main menu, choose the item: "Curves ⇒ Poverty Gap Quantile".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size Variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
Poverty line	z_1	z_2	Compulsory
p	p_1	p_2	Compulsory

Commands:

- **"Crossing"** : to search the first intersection of the curves. If the two curves intersect, DAD indicates the co-ordinates of the first intersection and their standard deviation if the option of computing with standard deviation is chosen. To seek an intersection over a particular range, use **"Range"**
- **"Difference" :** to compute the difference $g_1(z_1;p_1)-g_2(z_2;p_2)$.
- **"Graph" :** to draw the difference $g_1(z_1,p)-g_1(z_1;p)$ as a function of p.
- **"Range":** to specify the range for the search for a crossing between the two curves. This also specifies the range of the horizontal axis.

Lorenz curve and generalised Lorenz curve

The Lorenz curve at p for a population subgroup k is given by:

$$L(k;p)=\frac{\sum_{i=1}^{n} w_i^k y_i I(y_i \le Q(k;p))}{\sum_{i=1}^{n} w_i^k y_i}$$

where $I(y_i \le Q(k;p))=1$ if $y_i \le Q(k;p)$ and 0 otherwise. $Q(k;p)$ is the p-quantile of the subgroup k.

The generalised Lorenz curve at p for a population subgroup k is:

$$GL(k;p)=\mu.L(k;p)$$

Remark: The application for the Lorenz curve is similar in structure to the one for the generalised Lorenz curve

Case 1: One distribution

To compute the Lorenz curve for one distribution:

1- From the main menu, choose the item: "Curves ⇒ Lorenz curve".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size Variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
rho	ρ	Compulsory
p	p	Compulsory

Commands:

- "**Compute**": to compute $L(k;p)$. To compute the standard deviation, choose the option for computing with standard deviation.
- "**Graph**": to draw the Lorenz curve. To specify a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.
- "**Range**": to specify the range of the horizontal axis.

To compute the standard deviation, choose the option for computing with standard deviation.

Case 2 : Two distributions

To compute the Lorenz curve with two distributions:

1- From the main menu, choose the item: "Curves ⇒ Lorenz curve".
2- In the configuration of application, choose 2 for the number of distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Size Variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
rho	ρ_1	ρ_2	Compulsory
p	p_1	p_2	Compulsory

Commands:

- **"Crossing"**: to search the first intersection of the curves. If the two curves intersect, DAD indicates the co-ordinates of the first intersection and their standard deviation if the option of computing with standard deviation is chosen. To seek an intersection over a particular range, use **"Range".**
- **"Difference":** to compute the difference: $L_1(k_1;p_1) - L_1(k_{2;}p_2)$.
- **"Graph":** to draw the difference $L_1(k_1;p) - L_2(k_2;p)$ as a function of p.
- **"Range":** to specify the range for the search of a crossing between the two curves. This also specifies the range of the horizontal axis.
- **"S-Gini":** to compute the difference $I_1(k_1;\rho) - I_2(k_2;\rho)$.
- "**Covariance":** to compute the following covariance matrix:

$$\begin{matrix} \mathrm{Cov}(L_1(k_1;0.1),L_2(k_2;0.1)) & \mathrm{Cov}(L_1(k_1;0.1),L_2(k_2;0.2)) & \cdots & \mathrm{Cov}(L_1(k_1;0.1),L_2(k_2;1)) \\ \mathrm{Cov}(L_1(k_1;0.2),L_2(k_2;0.1)) & \mathrm{Cov}(L_1(k_1;0.2),L_2(k_2;0.2)) & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(L_1(k_1;1),L_2(k_2;0.1)) & \mathrm{Cov}(L_1(k_1;1),L_2(k_2;0.2)) & \cdots & \mathrm{Cov}(L_1(k_1;1),L_2(k_2;1)) \end{matrix}$$

Concentration curve and generalised concentration curve

The concentration curve for the variable T ordered in terms of y at p and for a population subgroup k is:

$$C_T(k;p) = \frac{\sum_{i=1}^{n} w_i^k T_i I(y_i \le Q(k;p))}{\sum_{i=1}^{n} w_i^k T_i}$$

where $I(y_i \le Q(k;p)) = 1$ if $y_i \le Q(k;p)$ and 0 otherwise. $Q(k;p)$ is the p-quantile of *y* for the subgroup k.
The generalised concentration curve at p for a population subgroup p is:

$$C_T(k;p) = \frac{\sum_{i=1}^{n} w_i^k T_i I(y_i \le Q(k;p))}{\sum_{i=1}^{n} w_i^k}$$

Remark: The application for the concentration curve is similar in structure to the one for the generalised concentration curve

Case 1: One distribution

To compute the concentration curve for one distribution:

1- From the main menu, choose the item: "Curves ⇒ concentration curve".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	T	Compulsory
Ranking variable	y	Compulsory
Size Variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
rho	ρ	Compulsory
p	p	Compulsory

Commands:

- "**Compute"**: to compute the concentration curve $C(k;p)$. To compute the standard deviation, choose the option for computing with standard deviation.
- "**Graph"**: to draw the concentration curve. To specify a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.
- "**Range:** to specify the range of the horizontal axis.

- To compute the standard deviation, choose the option for computing with standard deviation.

Case 2: Two distributions

To compute the concentration curve of two distributions:

1- From the main menu, choose the item: "Curves ⇒ Concentration curve".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	**Distribution 2**	
Ranking variable	y^1	y^2	Compulsory
Variable of interest	T^1	T^2	Compulsory
Size Variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
rho	ρ_1	ρ_2	Compulsory
p	p_1	p_2	Compulsory

Commands:

- **"Crossing":** to search the first intersection of the curves. If the two curves intersect, DAD indicates the co-ordinates of the first intersection and their standard deviation if the option of computing with standard deviation is chosen. To seek an intersection over a particular range, use **"Range".**
- **"Difference":** to compute the difference in the concentration curves.
- **"Graph":** to draw the difference in the curves as a function of p.
- **"Range":** to specify the range for the search of a crossing between the two curves. This also specifies the range of the horizontal axis.
- **"S-Gini":** to compute the difference $IC_1(k_1;\rho) - IC_2(k_2;\rho)$.
- "**Covariance":** to compute the following covariance matrix:

$$\begin{matrix} \mathrm{Cov}(C_1(k_1;0.1), C_2(k_2;0.1)) & \mathrm{Cov}(C_1(k_1;0.1), C_2(k_2;0.2)) & \cdots & \mathrm{Cov}(C_1(k_1;0.1), C_2(k_2;1)) \\ \mathrm{Cov}(C_1(k_1;0.2), C_2(k_2;0.1)) & \mathrm{Cov}(C_1(k_1;0.2), C_2(k_2;0.2)) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(C_1(k_1;1), C_2(k_2;0.1)) & \mathrm{Cov}(C_1(k_1;1), C_2(k_2;0.2)) & \cdots & \mathrm{Cov}(C_1(k_1;1), C_2(k_2;1)) \end{matrix}$$

The Cumulative Poverty Gap (CPG) curve

The CPG curve at p for a subgroup k and poverty line z is:

$$G(k;p;z) = \frac{\sum_{i=1}^{n} w_i^k (z - y_i)_+ I(y_i \le Q(k;p))}{\sum_{i=1}^{n} w_i^k}$$

Case 1: One distribution

To compute the CPG curve for one distribution:
1- From the main menu, choose the item: "Curves ⇒ CPG curve".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size Variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Poverty line	z	Compulsory
p	p	Compulsory

Commands:

- "**Compute**": to compute $G(k;p;z)$. To compute the standard deviation, choose the option for computing with standard deviation.
- "**Graph**": to draw the curve as a function according of p. To specify a range for the horizontal axis, choose the item " Graph Management ⇒ Change range of x " from the main menu.
- To compute the standard deviation, choose the option for computing with standard deviation.

Case 2: Two distributions

To reach the application for two distributions:

1- From the main menu, choose the item: "Curves ⇒ CPG curve".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	**Distribution 2**	
Variable of interest	y^1	y^2	Compulsory
Size Variable	s^1	s^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
Poverty line	z_1	z_2	Compulsory
rho	ρ_1	ρ_2	Compulsory
p	p_1	p_2	Compulsory

Commands:

- **"Crossing"**: to search the first intersection of the curves. If the two curves intersect, DAD indicates the co-ordinates of the first intersection and their standard deviation if the option of computing with standard deviation is chosen. To seek an intersection over a particular range, use **"Range".**
- **"Difference":** to compute the difference: $G_1(k_{1;}p_1;z) - G_2(k_{2;}p_2;z)$.
- **"Graph":** to draw the difference $G_1(k_1;p;z_1) - G_2(k_2;p;z_2)$ as a function of p.
- **"Range":** to specify the range for the search for a crossing between the two curves. This also specifies the range of the horizontal axis.
- **"S-Gini":** to compute the difference $P_1(z_1;\rho) - P_2(z_1;\rho)$.
- "**Covariance":** to compute the following covariance matrix:

$$\begin{matrix} Cov(G_1(k_1;0.1;z_1),G_2(k_2;0.1;z_2)) & Cov(G_1(k_1;0.1;z_1),G_2(k_2;0.2;z_2)) & \cdots & Cov(G_1(k_1;0.1;z_1),G_2(k_2;1;z_2)) \\ Cov(G_1(k_1;0.2;z_1),G_2(k_2;0.1;z_2)) & Cov(G_1(k_1;0.2;z_1),G_2(k_2;0.2;z_2)) & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ Cov(G_1(k_1;1;z_1),G_2(k_2;0.1;z_2)) & Cov(G_1(k_1;1;z_1),G_2(k_2;0.2;z_2)) & \cdots & Cov(G_1(k_1;1;z_1),G_2(k_2;1;z_2)) \end{matrix}$$

C-Dominance Curve

The j^{th} **C**ommodity or **C**omponent dominance curve is defined as follows:

$$CD^j(k;z,s)=\begin{cases}(s-1)\dfrac{1}{\sum_{i=1}^{n} w_i^k}\sum_{i=1}^{n} w_i^k (z-y_i)_+^{s-2} y_i^j & \text{if} \quad s\geq 2 \\ E\left[y^j \mid y=z\right] f(z)=\dfrac{\sum_{i=1}^{n} w_i^k K(z-y_i)_+ y_i^j}{\sum_{i=1}^{n} w_i^k} & \text{if} \quad s=1\end{cases}$$

where K( ) is a kernel function. Dominance of order s is checked by setting α=s-1.

The C-Dominance curve normalized by z, which is denoted by $\overline{CD}$, is given by:

$$\overline{CD}^j(k;z,s)=\begin{cases}\dfrac{(s-1)}{z^{\alpha}}\dfrac{1}{\sum_{i=1}^{n} w_i^k}\sum_{i=1}^{n} w_i^k (z-y_i)_+^{s-2} y_i^j & \text{if} \quad s\geq 2 \\ E\left[y^j \mid y=z\right] f(z)=\dfrac{\sum_{i=1}^{n} w_i^k K(z-y_i)_+ y_i^j}{\sum_{i=1}^{n} w_i^k} & \text{if} \quad s=1\end{cases}$$

The C-Dominance curve normalized by the mean is defined as $\dfrac{CD^j}{\mu^j}$, and the C-Dominance curve normalized both by z and the mean equals: $\dfrac{\overline{CD}^j}{\mu^j}$.

Case 1: One distribution

To compute the C-Dominance curve for one distribution:

1- From the main menu, choose: "Curves ⇒ C-Dominance curve".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Component	y^j	Compulsory
Size Variable	sz	Optional
Group Variable	c	Optional
Group Number	k	Optional
Order s	s	Compulsory
Poverty line	z	Compulsory

Among the buttons, you will find:

- "**Compute":** to compute the C-Dominance curve at z and for a given alpha. To obtain the standard deviation, choose the option for computing with a standard deviation.
- **"Graph":** to draw the value of the C-Dominance curve over a range of z.

Case 2: Two distributions

To reach the application for two distributions:

1- From the main menu, choose: "Curves ⇒ C-Dominance curve ".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	Vectors or parameters		Choice is:
	Distribution 1	Distribution 2	
Variable of interest	y^1	y^2	Compulsory
Component	$y^{1,j}$	$y^{2,j}$	Compulsory
Size Variable	sz^1	sz^2	Optional
Group Variable	c^1	c^2	Optional
Group Number	k_1	k_2	Optional
Poverty line	z_1	z_2	Compulsory
Order s	s_1	s_2	Compulsory

Commands:

- **"Difference":** to compute the difference: $CD^{1,j}(k;z,s) - CD^{2,j}(k;z,s)$.
- **"Graph":** to draw the difference $CD^{1,j}(k;z,s) - CD^{2,j}(k;z,s)$ as a function of z.
- **"Range":** to specify the range of the horizontal axis.

Redistribution

This section regroups the following applications:

1- Estimating the progressivity of a tax or a transfer.
2- Comparing the progressivity of two taxes or two transfers.
3- Comparing the progressivity of a transfer and a tax.
4- Estimating horizontal inequity.
5- Estimating redistribution.
6- Estimating a coefficient of concentration.

Estimating the progressivity of a tax or a transfer

Let:
- X be gross income;
- T be a tax;
- B be a transfer.

1) TR progressivity:

A tax T is TR-progressive if $L_X(p) - C_T(p) > 0 \quad \forall p \in]0,1[$

A transfer B is TR-progressive if $C_B(p) - L_X(p) > 0 \quad \forall p \in]0,1[$

2) IR-progressivity:

A tax T is IR-progressive if $C_{X-T}(p) - L_X(p) > 0 \quad \forall p \in]0,1[$

A transfer B is IR-progressive if $C_{X+B}(p) - L_X(p) > 0 \quad \forall p \in]0,1[$

To reach this application:

1- From the main menu, choose the item: «Redistribution ⇒ Tax or transfer".
2- Specify if you wish to estimate the progressivity of a tax or of a transfer.
3- Choose the approach to be either TR or IR.
4- Choose the different vectors and parameter values as follows

Indication	Variables or parameters	Choice is:
Gross income	X	Compulsory
Tax (transfer)	T or B	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
rho	ρ	Compulsory
p	p	Compulsory

Commands:

- The command "**S-Gini**": to compute:

	TR Approach	IR Approach
Tax	$IC_T(\rho) - I_X(\rho)$	$I_X(\rho) - IC_{X-T}(\rho)$
Transfer	$I_X(\rho) - IC_B(\rho)$	$I_X(\rho) - IC_{X+B}(\rho)$

where $IC(\rho)$ is the S-Gini coefficient of concentration and $I(\rho)$ is the S-Gini index of inequality.

- The command "**Crossing**": to seek the first intersection of the concentration and Lorenz curves. DAD indicates the co-ordinates of that first intersection and their standard deviation if the option of computing with standard deviation is chosen.
- The command "**Difference**": to compute:

	TR Approach	IR Approach
Tax	$L_X(p) - C_T(p)$	$C_{X-T}(p) - L_X(p)$
Transfer	$C_B(p) - L_X(p)$	$C_{X+B}(p) - L_X(p)$

- The command "**Range":** to specify a range of p for the search of the first intersection between the two curves. The command also allows to specify the range of the horizontal axis in the drawing of a graph.
- The command "**Graph":** to draw the following differences as a function of p:

	TR Approach	IR Approach
Tax	$L_X(p) - C_T(p)$	$C_{X-T}(p) - L_X(p)$
Transfer	$C_B(p) - L_X(p)$	$C_{X+B}(p) - L_X(p)$

Comparing the progressivity of two taxes or transfers

- Let:
- X be gross income;
- $T1$ and $T2$ be two taxes;
- $B1$ and $B2$ be two transfers.

1) TR Approach :

$T1$ is more TR-progressive than $T2$ if : $C_{T2}(p) - C_{T1}(p) > 0 \quad \forall p \in]0,1[$

$B1$ is more TR-progressive than $B2$ if : $C_{B1}(p) - C_{B2}(p) > 0 \quad \forall p \in]0,1[$

2) IR approach :

$T1$ is more IR-progressive than $T2$ if : $C_{X-T1}(p) - C_{X-T2}(p) > 0 \quad \forall p \in]0,1[$

$B1$ is more IR-progressive than $B2$ if : $C_{X+B1}(p) - C_{X+B2}(p) > 0 \quad \forall p \in]0,1[$

To reach this application:

1- From the main menu, choose the item: «Redistribution ⇒ Transfer-Tax vs Transfer-Tax".
2- In front of the indicators "**Tax (Transfer)**" 1 and 2, specify the two vectors of taxes or transfers.
3- Choose the approach to be either TR or IR.
4- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Gross income	X	Compulsory
Tax (transfer) 1	T1 or B1	Compulsory
Tax (transfer) 2	T2 or B2	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
rho	ρ	Compulsory
p	p	Compulsory

Commands:

- The command "**S-Gini**": to compute:

	TR Approach	IR Approach
Tax	$IC_{T1}(\rho) - IC_{T2}(\rho)$	$IC_{X-T2}(\rho) - IC_{X-T1}(\rho)$
Transfer	$IC_{B2}(\rho) - IC_{B1}(\rho)$	$IC_{X+B2}(\rho) - IC_{X+B1}(\rho)$

where $IC(\rho)$ is the S-Gini coefficient of concentration.

- The command "**Crossing**": to seek the first intersection of the two concentration curves. DAD indicates the co-ordinates of that first intersection and their standard deviation if the option of computing with standard deviation is chosen.
- The command "**Difference**": to compute:

	TR Approach	IR Approach
Tax	$C_{T2}(p) - C_{T1}(p)$	$C_{X-T1}(p) - C_{X-T2}(p)$
Transfer	$C_{B1}(p) - C_{B2}(p)$	$C_{X+B1}(p) - C_{X+B2}(p)$

- The command "**Range**"**:** to specify a range of *p* for the search of the first intersection between the two curves. The command also allows to specify the range of the horizontal axis in the drawing of a graph.
- The command "**Graph**"**:** To draw the following curves as a function of *p*:

	TR Approach	IR Approach
Tax	$C_{T2}(p) - C_{T1}(p)$	$C_{X-T1}(p) - C_{X-T2}(p)$
Transfer	$C_{B1}(p) - C_{B2}(p)$	$C_{X+B1}(p) - C_{X+B2}(p)$

Comparing the progressivity of a transfer and of a tax

Let :

- X be gross income;
- T be a tax;
- B a transfer.

TR Approach:

The transfer B is more TR-progressive than a tax T if: $C_B(p) - L_X(p) > L_X(p) - C_T(p)$

$\forall p \in]0,1[$

IR Approach :

The transfer B is more IR-progressive than a tax T if: $C_{X+B}(p) > C_{X-T}(p)$

$\forall p \in]0,1[$

To reach this application:

1- From the main menu, choose the item: «Redistribution ⇒ Transfer vs Tax".
2- Choose the approach to be either TR or IR
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Gross income	X	Compulsory
Variable of tax	T	Compulsory
Variable of transfer	B	Compulsory
Size variable	s	Optional
Group variable	c	Optional
Group number	k	Optional
Rho	ρ	Compulsory
p	p	Compulsory

Commands:

- The command "**S-Gini**": to compute:

TR Approach	IR Approach
$2I_X(\rho) - IC_T(\rho) - IC_B(\rho)$	$IC_{X-T}(\rho) - IC_{X+B}(\rho)$

where $IC(\rho)$ is the coefficient of concentration.

- The command "**Crossing**" : to seek the first point at which the progressivity ranking of the tax and transfer is reversed. DAD indicates the co-ordinates of that first reversal and their standard deviation if the option of computing with standard deviation is chosen. These co-ordinates are:

TR Approach	IR Approach
$C_B(p) - L_X(p)$	$C_{X+B}(p)$

- The command "**Difference**" : to compute:

TR Approach	IR Approach
$C_T(p) + C_B(p) - 2L_X(p)$	$C_{X+B}(p) - C_{X-T}(p)$

- The command "**Range":** to specify a range of *p* for the search of the first reversal of the progressivity ranking. The command also allows to specify the range of the horizontal axis in the drawing of a graph.
- The command "**Graph :** to draw the following curves as a function of p:

TR Approach	IR Approach
$C_T(p)+C_B(p)-2L_X(p)$	$C_{X+B}(p)-C_{X-T}(p)$

Horizontal inequity

A tax or a transfer T causes reranking (and is therefore horizontally inequitable) if:

Tax : $C_{X-T}(p)-L_{X-T}(p)>0$ for at least one value of $p\in\left]0,1\right[$
Transfer : $C_{X+T}(p)-L_{X+T}(p)>0$ for at least one value of $p\in\left]0,1\right[$

To reach this application:

1- From the main menu, choose the item: «Redistribution ⇒ Horizontal inequity".
2- Specify if you are using a tax or a transfer.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Gross income	X	Compulsory
Tax (transfer)	T or B	Compulsory
Size variable	s	Optional
Group variable	c	Optional
Group numberof interest	k	Optional
rho	ρ	Compulsory
p	p	Compulsory

Commands:

- The command "**S-Gini**" : to compute:

Tax	Transfer
$I_{X-T}(\rho)-IC_{X-T}(\rho)$	$I_{X+B}(\rho)-IC_{X+B}(\rho)$

- The command "**Difference**" : to compute:

Tax	Transfer
$C_{X-T}(p)-L_{X-T}(p)$	$C_{X+B}(p)-L_{X+B}(p)$

- The command "**Range":** to specify the range of the horizontal axis in the drawing of a graph.

- The command "**Graph" :** To draw the following curves as a function of *p*:

Tax	Transfer
$C_{X-T}(p) - L_{X-T}(p)$	$C_{X+B}(p) - L_{X+B}(p)$

Redistribution

A tax or a transfer T redistributes if :

Tax : $L_{X-T}(p) - L_X(p) > 0 \qquad \forall p \in]0,1[$
Transfer : $L_{X+B}(p) - L_X(p) > 0 \qquad \forall p \in]0,1[$

To reach this application:

1- From the main menu, choose the item: «Redistribution ⇒ Redistribution".
2- Specify if you are using a tax or a transfer.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Basic variable	X	Compulsory
Interest variable	T or B	Compulsory
Size variable	s	Optional
Group variable	c	Optional
Group number	k	Optional
rho	ρ	Compulsory
p	p	Compulsory

Commands:

- The command "**S-Gini**": to compute:

Tax	Transfer
$I_X(\rho) - I_{X-T}(\rho)$	$I_X(\rho) - I_{X+B}(\rho)$

- The command "**Crossing**": to seek the first point at which the curves $L_{X-T}(p)$ and $L_X(p)$, or $L_{X+B}(p)$ and $L_X(p)$, cross. DAD indicates the co-ordinates of that first crossing and their standard deviation if the option of computing with standard deviation is chosen.
- The command "**Difference**: with this command, to compute:

Tax	Transfer
$L_{X-T}(p)-L_X(p)$	$L_{X+B}(p)-L_X(p)$

- The command "**Range":** to specify a range of p for the search of the first intersection between the two curves. The command also allows to specify the range of the horizontal axis in the drawing of a graph.
- The command "**Graph" :** to draw the following curves as a function of p:

Tax	Transfer
$L_{X-T}(p)-L_X(p)$	$L_{X+B}(p)-L_X(p)$

The coefficient of concentration

Let a sample contain n joint observations, (y_i, T_i), on a variable y and a variable T. Let observations be ordered in increasing values of y, in such a way that $y_i \le y_{i+1}$. The S-Gini coefficient of concentration of T for the group k is denoted as $IC_T(k;\rho)$ and defined as:

$$IC_T(k;\rho) = 1 - \frac{\sum_{i=1}^{n} \left[\frac{(V_i)^\rho - (V_{i+1})^\rho}{[V_1]^\rho} \right] T_i}{\mu_T} \quad \text{where } V_i = \sum_{h=i}^{n} w_h^k .$$

One distribution

To compute the coefficient of concentration for only one distribution:

1- From the main menu, choose the following item: "Redistribution ⇒ Coefficient of concentration".
2- In the configuration of the application, choose 1 distribution.
3- After confirming the configuration, the application appears. Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Ranking variable	y	Compulsory
Variable of interest	T	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group number	k	Optional
rho	ρ	Compulsory

Commands:

- The command "**Compute"**: to compute the coefficient of concentration. To compute the standard deviation of this index, choose the option for computing with standard deviation.
- The command "**Graph"**: to draw the value of the coefficient as a function of the parameter ρ. To specify a range for the horizontal axis, choose the item " Graph management $\Rightarrow$ Change range of x " from the main menu.

Two distributions

To reach this application:

1- From the main menu, choose the item: "Redistribution $\Rightarrow$ Coefficient of concentration".
2- In the configuration of application, choose 2 distributions.
3- Choose the different vectors and parameter values as follows:

Indication	**Vectors or parameters**		**Choice is:**
	Distribution 1	**Distribution 2**	
Variable of interest	T^1	T^2	Compulsory
Ranking variable	y^1	y^2	Compulsory
Size variable	s^1	s^2	Optional
Group variable	c^1	c^2	Optional
Group number	k_1	k_2	Optional
rho	ρ_1	ρ_2	Compulsory

Press « **Compute** » to compute the concentration coefficients and their difference for each of the two variables of interest. To compute the standard deviation of those estimates, choose the option for computing with standard deviation.

Distribution

Descriptive statistics

This application provides basic descriptive statistics on variables in the database: the mean, the standard deviation, and the minimum and the maximum values of each of the vectors.

To reach this application:

1- From the main menu, choose: "Distribution ⇒ Statistics".
2- Choose the data bases if you have activated two databases.
3- Choose the weight variable if the observations must be weighted.
4- Choose the group variable and the group number if you would like to compute the statistics for a specific group.

The results are as follows:

Name of variable 1	Mean	Standard deviation	Minimum	Maximum
Name of variable 2	Mean	Standard deviation	Minimum	Maximum
:	:	:	:	:

Statistics

This application computes basic descriptive statistics for a given variable of interest, as well as the ratio of two such variables. The application also computes the effect of the sampling design on the sampling error of these basic statistics.

1- Total $= \sum_i w_i \; x_i$

2- Mean $= \dfrac{\sum_i w_i x_i}{\sum_i w_i}$

3- Ratio $= \dfrac{\sum_i w_i x_i}{\sum_i w_i \; y_i}$

To activate this application for one distribution, follow these steps:

1- From the main menu, choose: "Distribution ⇒ Statistics".
2- In the configuration of application, choose 1 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is
Variable of interest 1	x	Compulsory
Size Variable 1	s(x)	Optional
Variable of interest 2	y	Optional
Size Variable 2	s(y)	Optional
Group Variable	c	Optional
Group Number	k	Optional

To activate this application for one distribution, follow these steps:

1- From the main menu, choose the item: "Distribution ⇒ Statistics".
2- In the configuration of application, choose 2 distribution.
3- Choose the different vectors and parameter values as follows:

Indication	Vector or parameter		Choice is
	Distribution 1	Distribution 2	
Variable of interest 1	x^1	x^2	Compulsory
Size Variable 1	$s(x)^1$	$s(x)^2$	Optional
Variable of interest 2	y^1	y^2	Optional
Size Variable 2	$s(y)^1$	$s(y)^2$	Optional
Group Variable	c^1	c^2	Optional
Group Number	k^1	k^2	Optional

Density function

The gaussian kernel estimator of a density function $f(x)$ is defined as:

$$\hat{f}(x) = \frac{\sum_i w_i K_i(x)}{\sum_{i=1}^{n} w_i} \quad \text{and} \quad K_i(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-0.5\,\lambda_i(x)^2\right) \quad \text{and} \quad \lambda_i(x) = \frac{x - x_i}{h}$$

where h is a bandwidth which acts as a “smoothing” parameter.

To reach this application:

1- From the main menu, choose the item: "Distribution ⇒ Density function".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Parameter	y	Compulsory
Smoothing parameter	h	Optional

On the first execution bar, you find:

- The command **"Compute":** to compute $f(x)$. To compute the standard deviation, choose the option for computing with standard deviation.
- The command **"Graph":** to draw the value of the function as a function of x. To specify a range for the horizontal axis, choose the item " Graph management ⇒ Change range of x " from the main menu.
- The command **"Range":** to specify the range of the horizontal axis

To compute the standard deviation, choose the option for computing with standard deviation.

Corrected boundary Kernel estimators

A problem occurs with kernel estimation when a variable of interest is bounded. It may be for instance that consumption is bounded between two bounds, a minimum and a maximum, and that we wish to estimate its density "close" to these two bounds. If the true value of the density at these two bounds is positive, usual kernel estimation of the density close to these two bounds will be biased. A similar problem occurs with non-parametric regressions. One way to alleviate these problems is to use a smooth "corrected" Kernel estimator, following a paper by Peter Bearse, Jose Canals and Paul Rilstone. A boundary-corrected Kernel density estimator can then be written as

$$\hat{f}(x) = \frac{\sum_i w_i K_i^*(x) K_i(x)}{\sum_{i=1}^{n} w_i}$$

where

$$K_i(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-0.5\,\lambda_i(x)^2\right) \quad \text{and} \quad \lambda_i(x) = \frac{x - x_i}{h}$$

and where the scalar $K_i^*(x)$ is defined as

$$K_i^*(x) = \psi(x)' P(\lambda_i(x))$$

$$P(\lambda) = \left(1 \quad \lambda \quad \frac{\lambda^2}{2!} \quad \cdots \quad \frac{\lambda^{s-1}}{(s-1)!}\right)$$

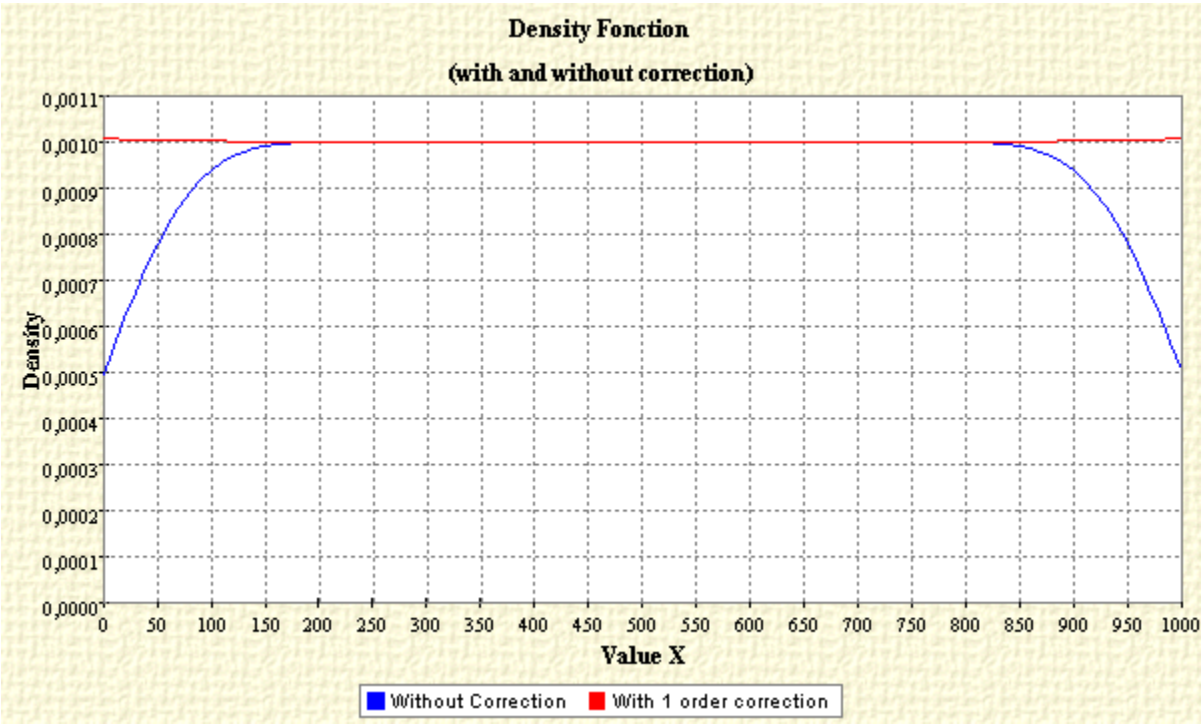
$$\psi(x) = M^{-1} l_s' = \left(\int_A^B K(\lambda)P(\lambda)P(\lambda)'d\lambda\right)^{-1} l_s \quad : A = \frac{x - max}{h}, \quad B = \frac{x - min}{h}, \quad l_s' = (1 \quad 0 \quad 0 \cdots 0)$$

min is the minimum bound, and *max* is the maximum one. *h* is the usual bandwidth. This correction removes bias to order h^s.

DAD offers four options, without correction, and with correction of order 1, 2 and 3.

Example 1:

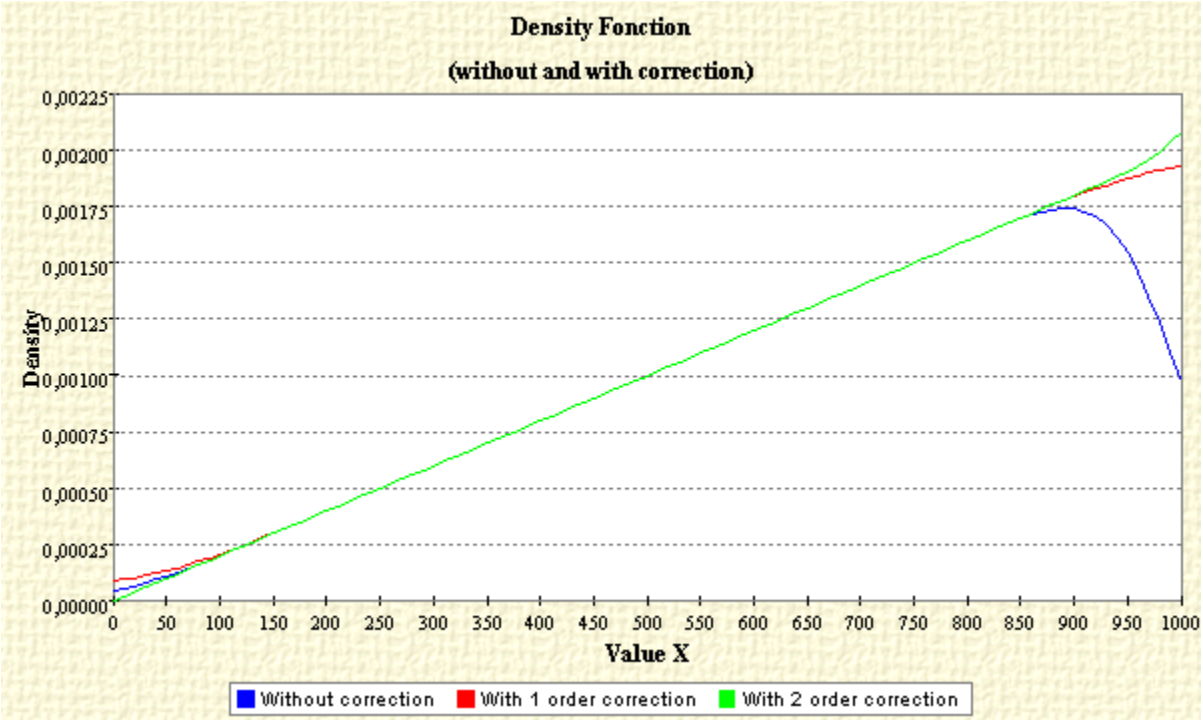
Suppose that an observed vector of interest y takes the form : y={1,2,3,…i+1….999,1000} because it is drawn from a uniform distribution. The density at any income between 0 and 1000 is the same and equals 1/1000. The following figure shows the impact of the above correction on the density estimation:



This shows that a correction of order 1 corrects well the boundary problem of estimating the density close to 0 and 1000.

Example 2:

Suppose that an observed vector of interest y takes the form : y={1,2,2,3,3,3,…,….1000,1000}. The total number of observations sums to N=1000*(1+1000)/2=50500. The population density equals f(x)=x/500.The following figure shows the impact of a correction of order 1 and 2 on the density estimation:



The joint density function

The gaussian kernel estimator of the joint density function $f(x,y)$ is defined as:

$$\hat{f}(x,y) = \frac{1}{\sum_{i=1}^{n} w_i h^2} \sum_{i=1}^{n} w_i \frac{1}{2.\pi} \exp\left(-\left(\frac{1}{2}\right)\left(\left(\frac{x-x_i}{h}\right)^2 + \left(\frac{y-y_i}{h}\right)^2\right)\right)$$

To reach this application:

1- From the main menu, choose the item: "Distribution ⇒ Joint density function".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	*x*	Compulsory
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	*c*	Optional
Group Number	k	Optional
Parameter	x	Compulsory
Parameter	*y*	Compulsory
Smoothing parameter	h	Optional

On the first execution bar, you find:

- The command **"Compute":** to compute the estimate of the joint density function. To compute the standard deviation, choose the option for computing with standard deviation

The distribution function

To reach this application:

1- From the main menu, choose the item: "Distribution ⇒ Distribution function".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Variable of interest	y	Compulsory
Size variable	s	Optional
Group Variable	c	Optional
Group Number	k	Optional
Parameter	y	Compulsory

On the first execution bar, you find:

- The command **"Compute":** to compute the estimate of the distribution function. To compute the standard deviation, choose the option for computing with standard deviation.
- The command **"Graph":** to draw the distribution function F(x) along values of x. To specify a range for the horizontal axis, choose the item " Graph management ⇒ Change range of x " from the main menu.
- The command **"Range":** to specify the range of the horizontal axis

Plot_Scatt_XY

- This application plots a scatter graph of two variables. To activate this application, choose from the main menu the item: "Distribution ⇒ Plot_Scatt_XY". When the window of this application appears, choose the two X and Y variables and click on the button "**Graph**". You can also use the command **"Range"** to specify the range of the horizontal axis (X).

Non-parametric regression and non-parametric derivative regression

The Gaussian kernel regression of y on x is as follows:

$$\Phi(y \mid x) = \frac{\alpha(x)}{\beta(x)} = \frac{\sum_i w_i K_i(x)\ y_i}{\sum_i w_i K_i(x)}$$

From this, the derivate of $\Phi(y \mid x)$ with respect to x is given by

$$\frac{\partial \Phi(y \mid x)}{\partial x} = \frac{\alpha(x)'}{\beta(x)} - \frac{\beta(x)'\alpha(x)}{\beta(x)^2}$$

Remark: the instructions for non-parametric derivative regression are similar to those for non-parametric regression

To reach this application:

1- From the main menu, choose the item: "Distribution ⇒ Non-parametric regression".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is:
Exogenous Variable (X)	x_i	Compulsory
Endogenous Variable (Y)	y_i	Compulsory
Size variable	s_i	Optional
Group Variable	c	Optional
Group Number	k	Optional
Level of (X) or (p)	x	Compulsory
Smoothing parameter	h	Optional

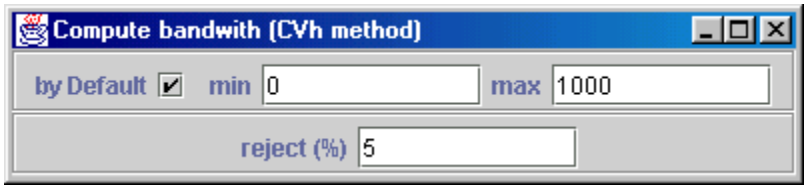
Remark 1: The option "Level" vs "Percentile" allows the estimation of the expected value of y either at a level of x or at a p-quantile for x.

Remark 2: The option "Normalised" vs "Not normalized" by the mean or by x allows the estimation of the expected value of y normalized or not by x or by the overall mean of y.

You will find:

- The command **"Compute":** to compute $\Phi(y \mid x)$. To compute its standard deviation, choose the option for computing with standard deviation.

- The command **"Compute h":** to compute an optimal bandwidth according to the cross-validation method of Härdle (1990), p. 159-160. When you click on this command, the following window appears, giving you the option of choosing the min/max bands and the percentage of observations to be rejected on each side of the range of x.



- The command **"Graph":** to draw $\Phi(y \mid x)$ as a function of x. To specify a range for the horizontal axis, choose the item " Graph management $\Rightarrow$ Change range of x " from the main menu.
- The command **"Range":** to specify the range of the horizontal axis.

Boundary-corrected non-parametric regression and non-parametric derivative regression

For the boundary-corrected non-parametric regression, the estimation is as follows:

$$\Phi(y \mid x) = \frac{\sum_i w_i K_i^*(x) K_i(x)\ y_i}{\sum_i w_i K_i^*(x) K_i(x)}$$

The boundary-corrected non-parametric derivate regression is obtained by differentiating the above with respect to x:

$$\Phi'(y \mid x) = \frac{\sum_i w_i \left(K_i^*(x)' K_i(x)\ y_i + K_i^*(x) K_i(x)'\ y_i\right)}{\sum_i w_i K_i^*(x) K_i(x)} - \frac{\sum_i w_i \left(K_i^*(x)' K_i(x) + K_i^*(x) K_i(x)'\right)}{\left(\sum_i w_i K_i^*(x) K_i(x)\right)^2}$$

Note that:

$$K_i^*(x) = \psi(x)' \, P(\lambda_i(x)) \text{ and } P(\lambda) = \left(1 \quad \lambda \quad \frac{\lambda^2}{2!} \quad \cdots \quad \frac{\lambda^{s-1}}{(s-1)!}\right)$$

$$\psi(x) = M^{-1} l_s{}' = \left(\int_A^B K(\lambda)P(\lambda)P(\lambda)'d\lambda\right)^{-1} l_s \quad : A = \frac{x - \max}{h}, \quad B = \frac{x - \min}{h}, \quad l_s{}' = (1 \quad 0 \quad 0 \cdots 0)$$

$$K_i^*(x)' = \frac{\partial M^{-1}(x)}{\partial x} l_s{}' \, P(w) + \frac{\partial P(\lambda(x))}{\partial x} M^{-1}(x) l_s{}' \text{ where}$$

$$\frac{\partial M^{-1}(x)}{\partial x} = -M^{-1}(x)\left[\frac{\partial M(x)}{\partial x}\right]M^{-1}(x)$$

Conditional standard deviation

A kernel estimator for the Conditional Standard Deviation of y at x can be defined as:

$$ST(x) = \left[\frac{\sum_i w_i K(x_i, x)(y_i - y(x))^2}{\sum_i w_i K(x_i, x)}\right]^{\frac{1}{2}}$$

where K is a kernel function and y(x) is the expected value of y conditional on x.
To reach this application:

1- From the main menu, choose: "Distribution ⇒ Conditional Standard Deviation".
2- Choose the different vectors and parameter values as follows:

Indication	Variables or parameters	Choice is
Exogenous Variable (X)	x_i	Compulsory
Endogenous Variable (Y)	y_i	Compulsory
Size variable	s_i	Optional
Group Variable	c	Optional
Group Number	k	Optional
Level of (X) or (p)		Compulsory
Smoothing parameter	h	Optional

Remark 1: The option "Level" vs "Percentile" allows the estimation of the conditional standard deviation of y either at a level of x or at a p-quantile for x.

You will find:

- The command **"Compute":** to compute ST(x).

- The command **"Graph":** to draw ST(x) as a function of x. To specify a range for the horizontal axis, choose the item " Graph management ⇒ Change range of x " from the main menu.
- The command **"Range":** to specify the range of the horizontal axis.

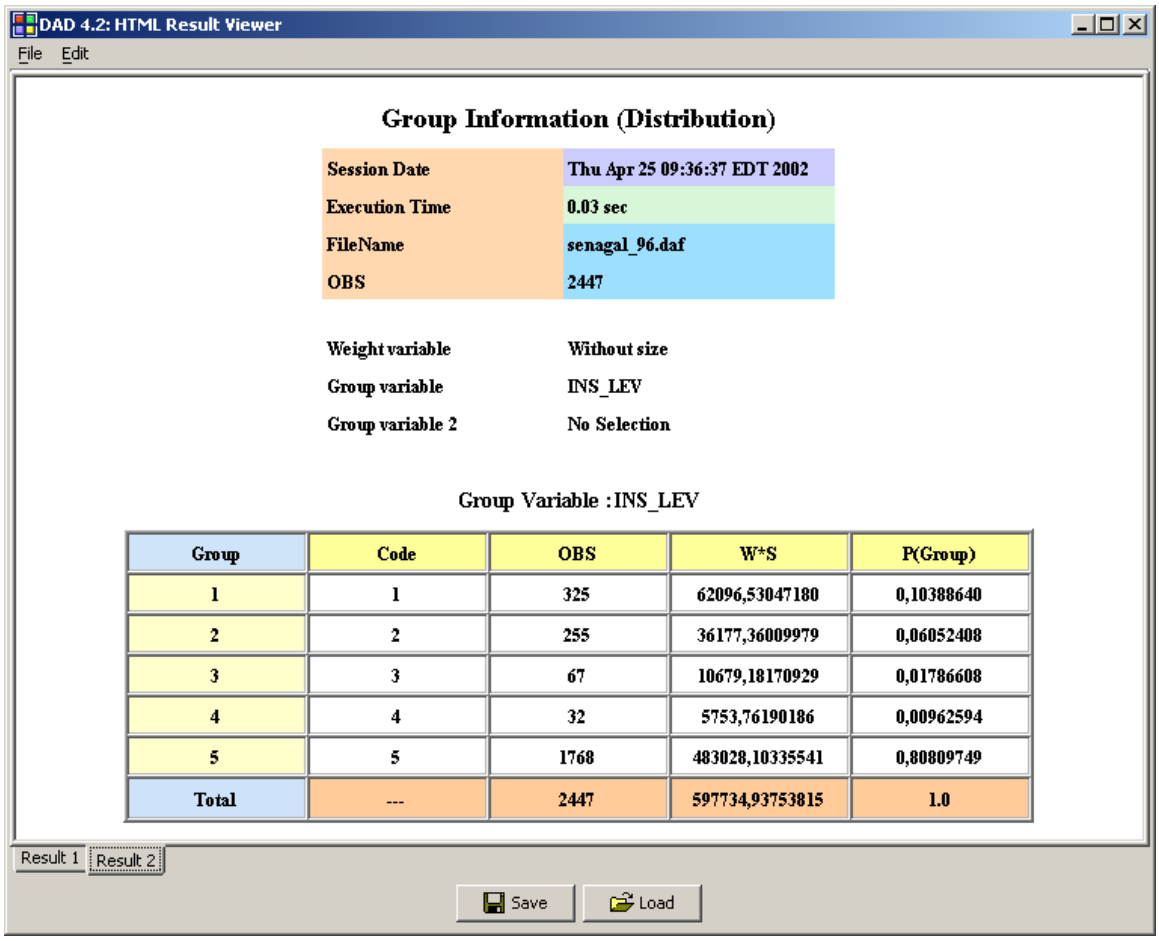
Group information

This application estimates the cross-group composition of a population. The group details are provided by the user through either or both of two *Group* variables.

To reach this application:

1- From the main menu, choose: "Distribution ⇒ Group Information".
2- Choose the first group variable.
3- Choose the size variable if the observations must be weighted by size.
4- Choose the second group variable if you would like cross-group (or cross-tabulation) information to be provided across two groups.

Example 1:


DAD 4.2: HTML Result Viewer
File Edit
Group Information (Distribution)
Session Date Thu Apr 25 09:36:37 EDT 2002
Execution Time 0.03 sec
FileName senagal_96.daf
OBS 2447
Weight variable Without size
Group variable INS_LEV
Group variable 2 No Selection
Group Variable : INS_LEV
Group Code OBS W*S P(Group)
1 1 325 62096,53047180 0,10388640
2 2 255 36177,36009979 0,06052408
3 3 67 10679,18170929 0,01786608
4 4 32 5753,76190186 0,00962594
5 5 1768 483028,10335541 0,80809749
Total --- 2447 597734,93753815 1.0
Result 1 Result 2
Save Load

This example uses only one group variable “INS-LEV” (level of instruction of the household head), categorized as

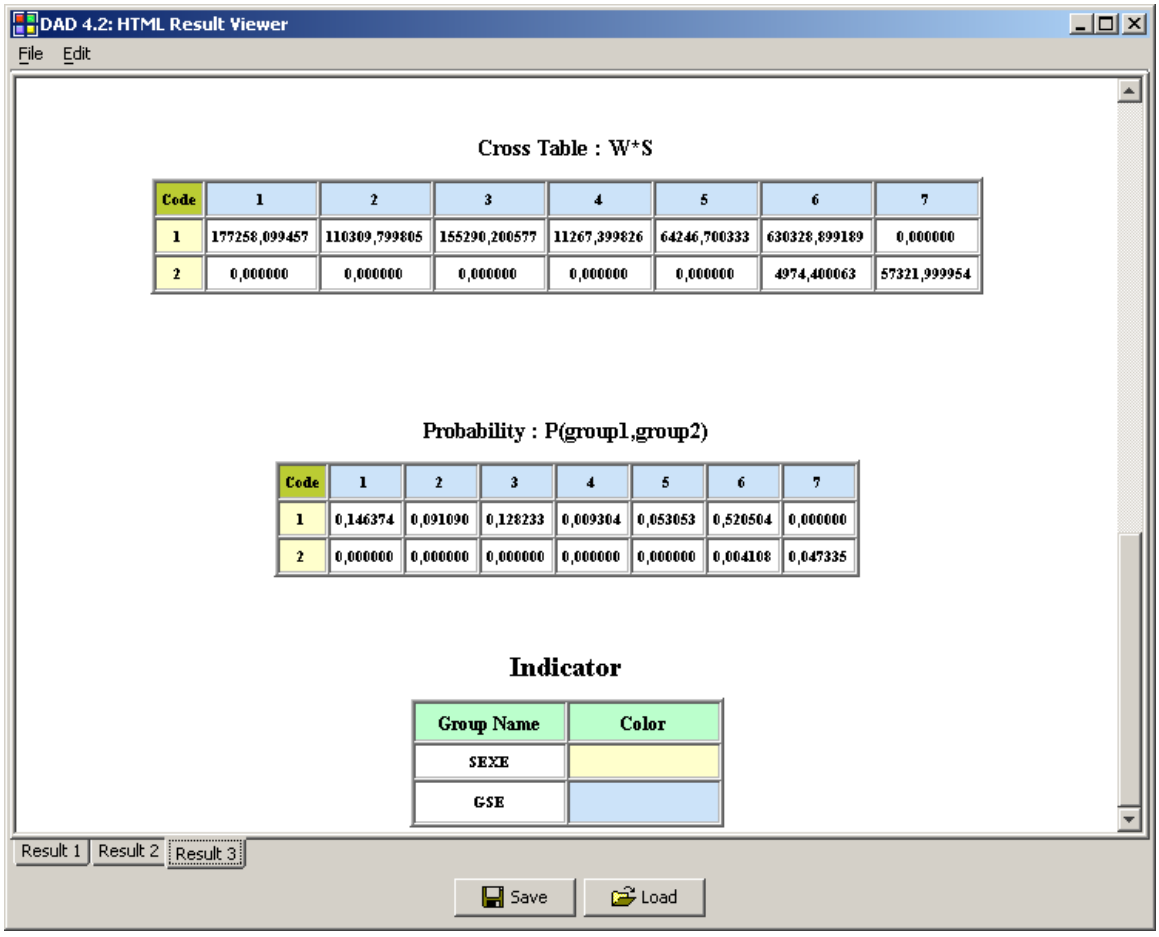
1- Primary
2- Secondary
3- Superior
4- Not available
5- None

The output shows:

Code	The exact code of the group
Group	The group number: (1,2,3,…)
OBS	The number of observations in the group
W*S	The sum of the products of Sampling Weight times Size
P(Group)	The estimated proportion of population found in that group

The use of two group variablesshows the following information:

Example 2:

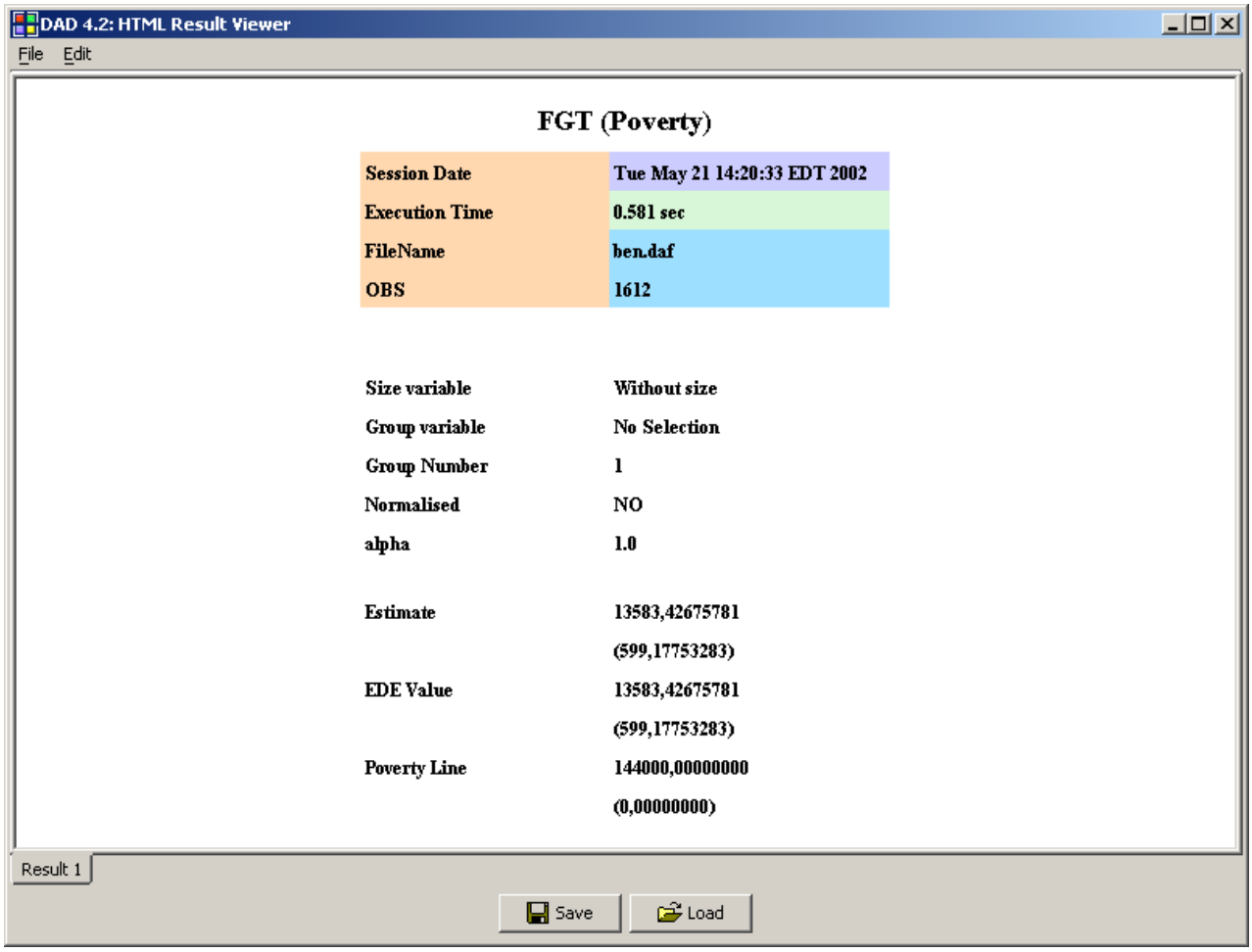


The *"Cross Table"* table shows the sum of the products of Sampling Weight times Size for those observations belonging to the two groups simultaneously. The second table, *"Probability"*, shows the estimated proportion of the population who belong to both of the groups.

The editing, saving and printing of results

Editing of results

Generally, the windows of results tack the following form:



The window contains the name of the application and the results of the execution. We can divide these results, displayed in the last figure, in three blocks:

1- General information: this first block is composed of:

Session date	Indicates the time at which the results were computed.
Execution time	Indicates the computation time.

2- The block of inputs composed by:

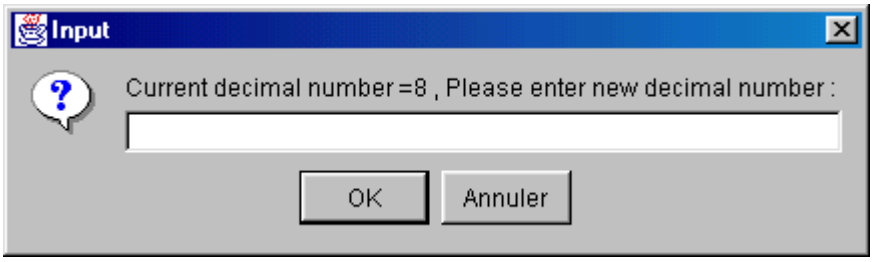
File name	indicates the name of the file that is used.
OBS	indicates the number of observations.
Parameter used	indicates the value of the parameter used for this computation

	(see also the illustrations for the computation of inequality indices).
Variable of interest	Indicates the name of the variable used to compute the index of inequality.
Size variable	indicates the size of variable.
Group variable	Indicates the vector that contains group indices (in this application, the choice of such a vector is optional)
Group Number	Indicates the selected group number (by default, its value equals one).
Parameter	Indicates to the user the names and the values of the parameters. The parameter names typically refer to the definition of indices and curves.
Options :	Indicates the options selected for this execution.

3- The third and last block contains the results of the execution.

Index value	Indicates the value of the index or point estimated. The value within parentheses indicates the standard deviation for this estimate.

One can select a number of decimal values for the printing of results. To do this, choose the command "Edit --> Change Decimal Number". The following window appears. Choose the desired number of decimals and confirm the choice by clicking on the button "OK"

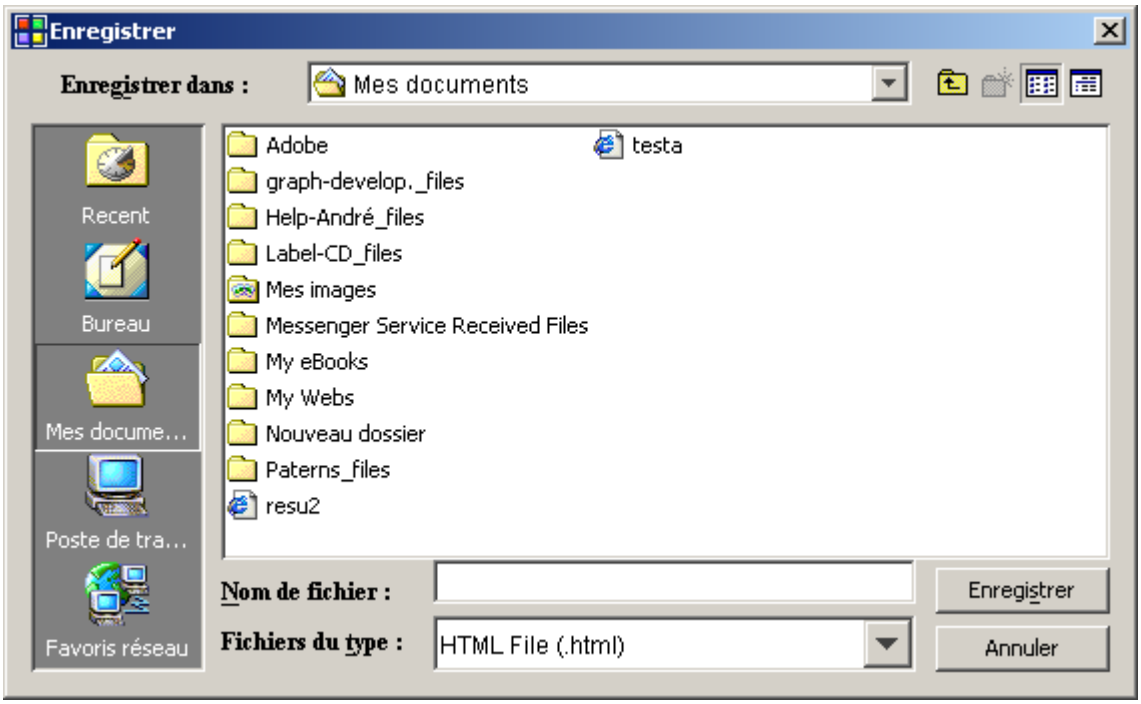


When another execution is performed, a new window appears with the information concerning this new execution. One can return to and edit the information on the previous executions by activating the window of the previous results. For this, click on the button representing the result (look on the bottom of the window for the buttons “Result1”, “Result2”.

Saving and printing results

DAD easily saves results in the HTML format. This allows the edition of these results with browsers like Explorer or Netscape.

To save the results, from the window of results choose the command “File -> Save (html format)”. The following window appears.



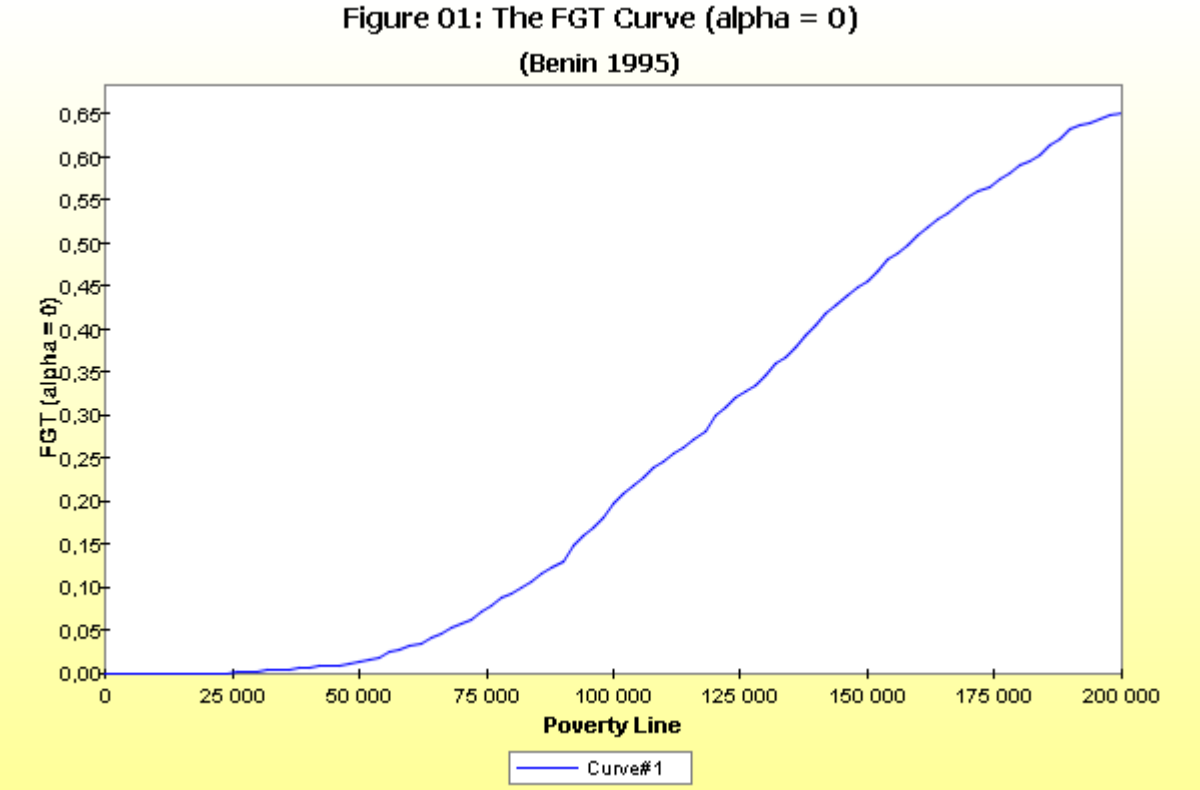
After making your choice of name and directory, click on the button **"Save"** to save the results.

To print these results, choose from the main window the command "File --> Print". The printing window appears; just choose the name of your printer and confirm by clicking on the button "OK".

Graphs in DAD4.2

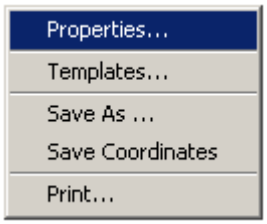
Drawing graphs

Most applications in DAD offer the possibility of plotting graphs to illustrate the results of those applications. For example, the FGT poverty index application can plot a curve of this index – against the Y axis – according to alternative levels of the poverty line – shown on the X axis – as in the following figure:



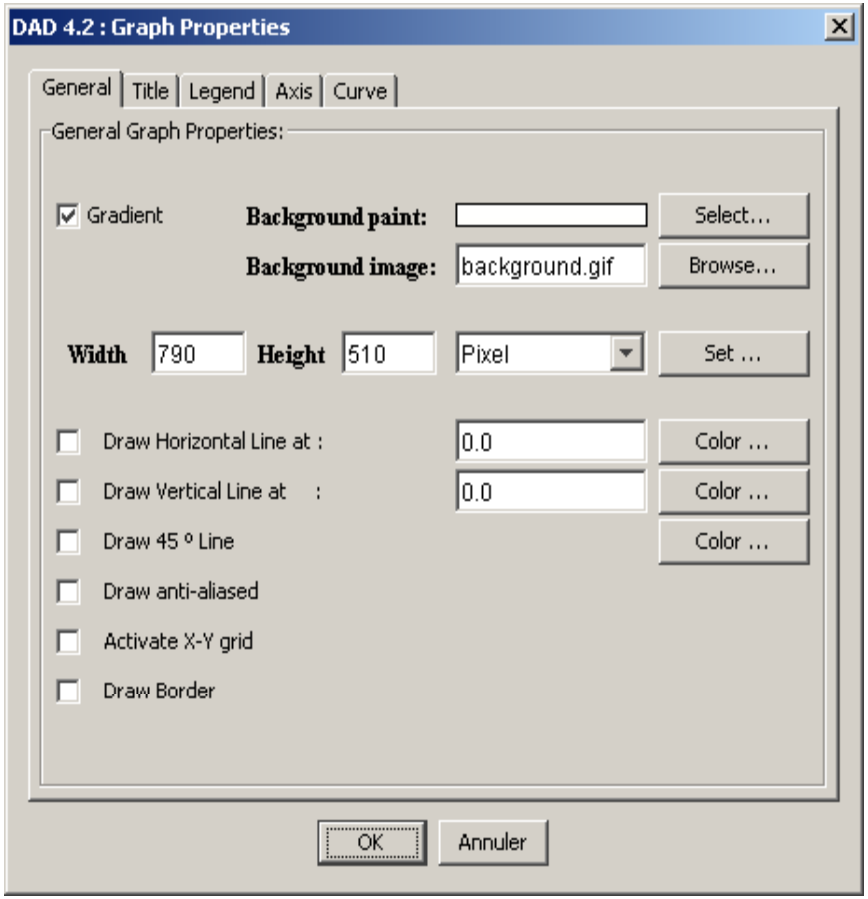
Changing graph properties

We can change many properties of a graph. For this, select the item: Tools⇒Properties. This can also be done by activating the Popup Menu. To activate the Popup Menu, click on the right button of the mouse when you are within the quadrant of graph. The items shows how to change graph properties in DAD.



The Popup Menu

General



Background paint: to select the background colour of the graph. We can also select the option "Gradient" for the background colour.

Background paint: to browse and select a picture (GIF or PNG) to be the background graph.

Width and Height: to indicate the desired width and height of the graph in pixels, inches or centimetres (click on the button **Set** to confirm your selection).

Draw Horizontal Line: to draw a horizontal line at a giving height of the Y-axis. Indicate that height and click the option.

Draw Vertical Line: to draw a vertical line at a giving value of the X-axis. Indicate that value and click the option.

Draw 45º Lines: to draw a 45º line.

Antia-aliasing option: One of the most important techniques in making graphics and text easy to read and pleasing to the eye on-screen is anti-aliasing. Anti-aliasing gets around the low 72dpi resolution of the computer monitor and makes objects appear smooth.

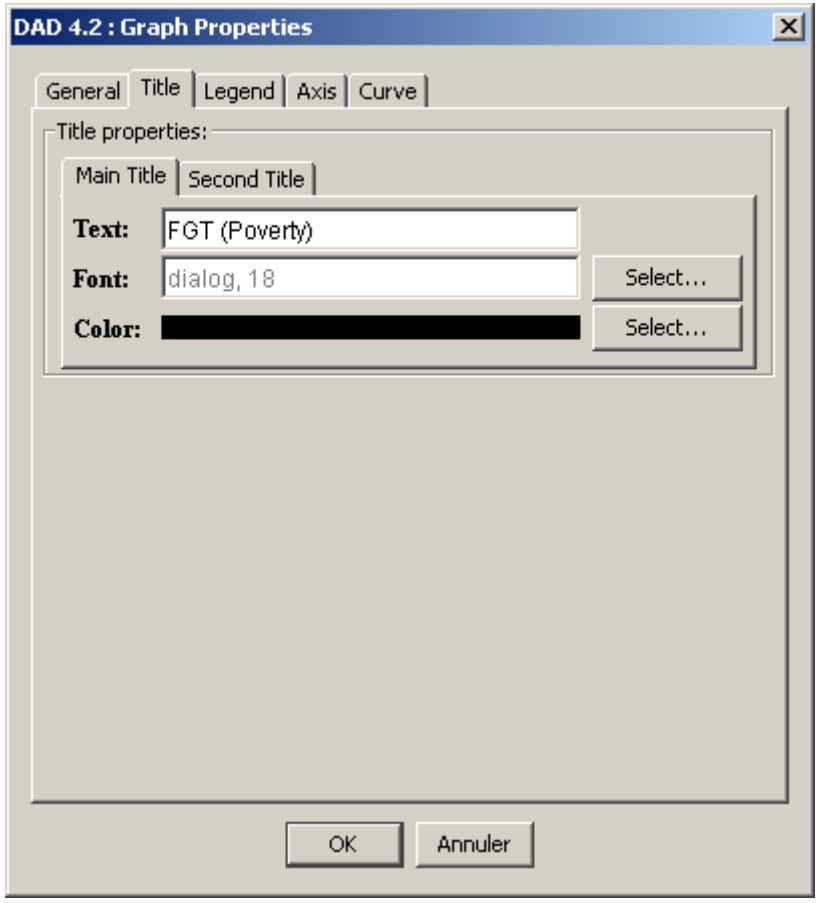
Activate X-Y grid: If this option is selected, a grid is plotted in the graph

Draw Border: If this option is selected, a border is plotted around the graph.

Title

Main Title: By default, the main title is the name of application. You can change the main title in the field **Text**. You can also change its font and its colour. To do this, just click on the button select and indicate the desired font or colour.

Second Title: By default, the second title is *Chart*. You can change or delete the second title in the field **Text**. You can also change its font and its colour. To do this, just click on the button select and indicate the desired font or colour.



Legend

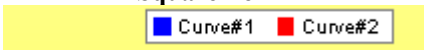
Background: to select the background colour of the legend quadrant.

Text font: to select the font of the text legends.

Text font: to select the colour of the text legends.

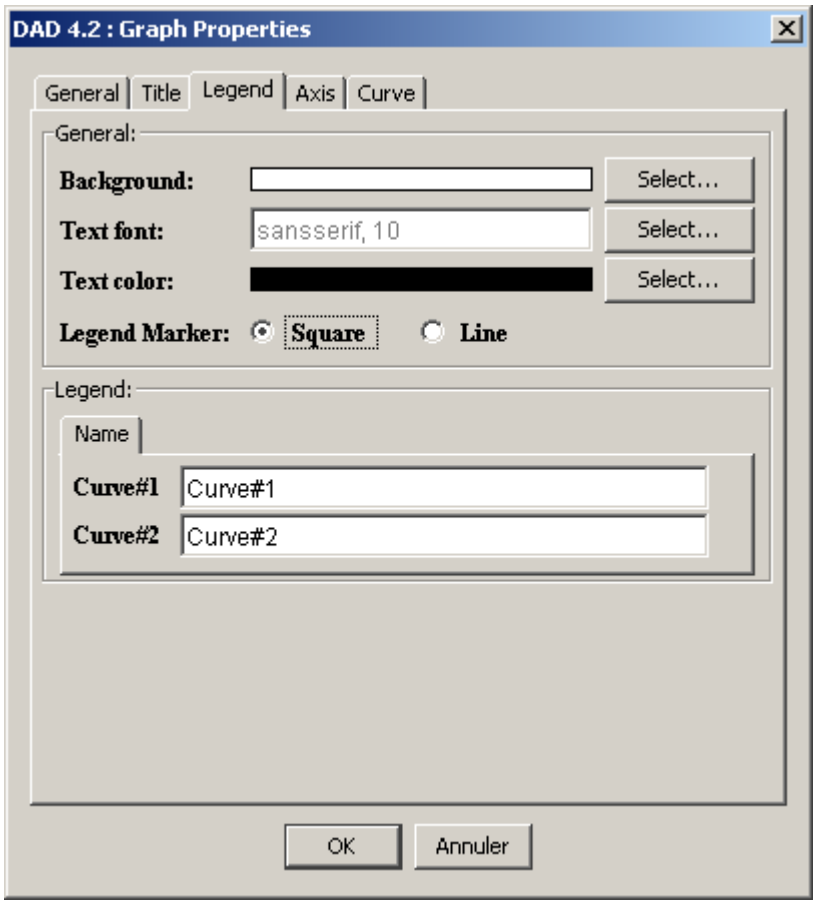
Legend Marker: to select Marker legends. By default, the markers have square form, but you can select the line form with this option.

Square Form



Line Form

Name: By default, the names of the curves are *curve#1, curve#2,etc.* You can change these names in these fields.



Axis

Remark: *The options for the horizontal axis are similar to those for the vertical axis.*

Name: By default, the name of the vertical axis is *Value Y.* You can change this name with this field.

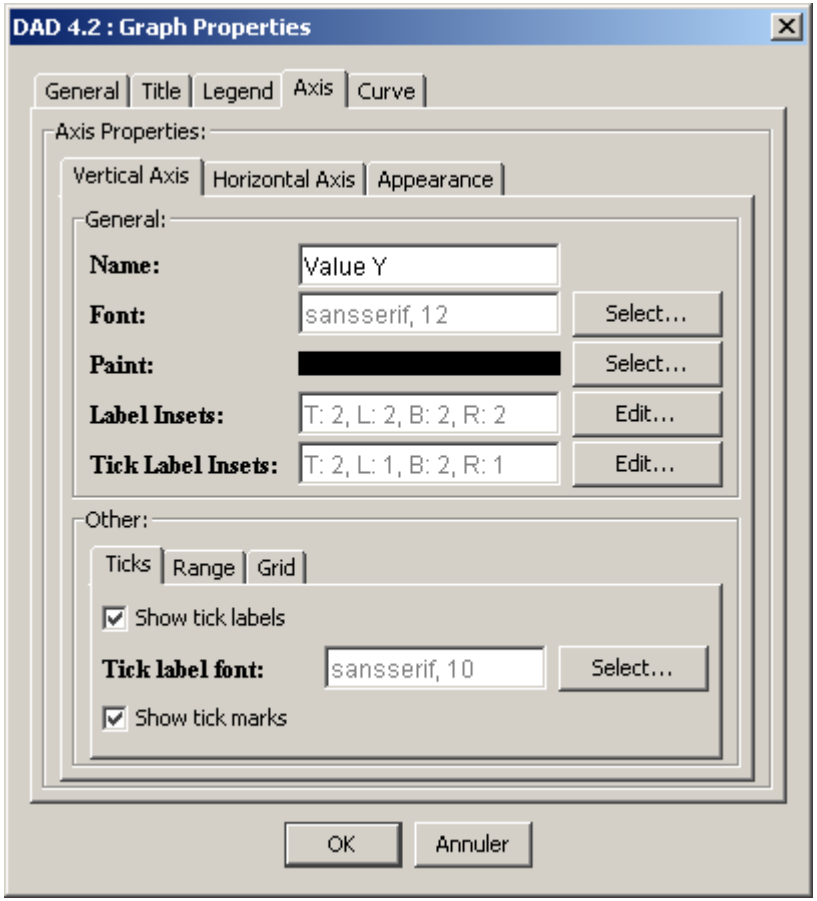
Font: to select the font of the name of the vertical axis.

Paint: to select the colour of the name of the vertical axis.

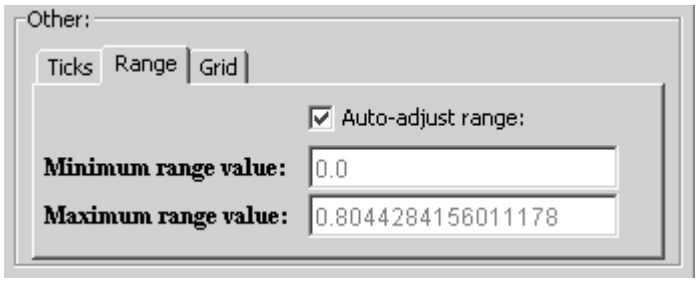
Label insets: to change the labels' position (*Top, Left, Bottom, Right)* indicated in pixels

Tick Label Insets: to change the Tick label position (*Top, Left, Bottom, Right)* indicated in pixels

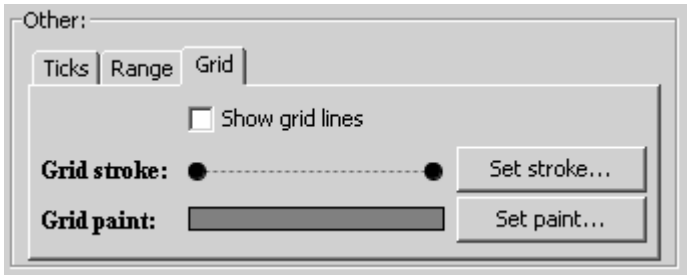
Other-Tick: to show or not to show the tick labels or the tick markers. You can also select the font of the tick labels.



Other-Range: to select the minimum and maximum values for the range of the vertical axis. To do this, unselect the option *Auto-adjust range*



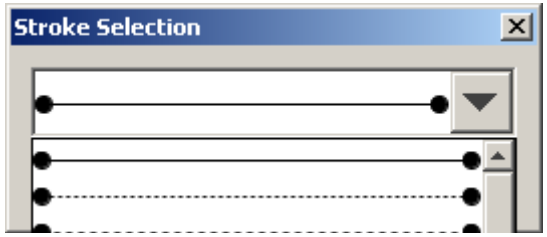
Other-Grid: To plot the horizontal grid lines, select the option *Show grid lines*. You can also select the stroke and the colour of these grid lines.



Curve

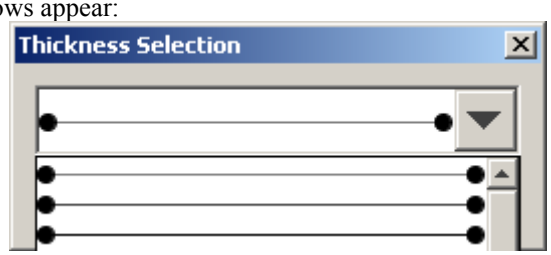
For every curve, a combination of the three flowing options can be chosen:

Curve Stroke: To choose the stroke of a giving curve, click on the button **Set stroke**. The following widows appear:

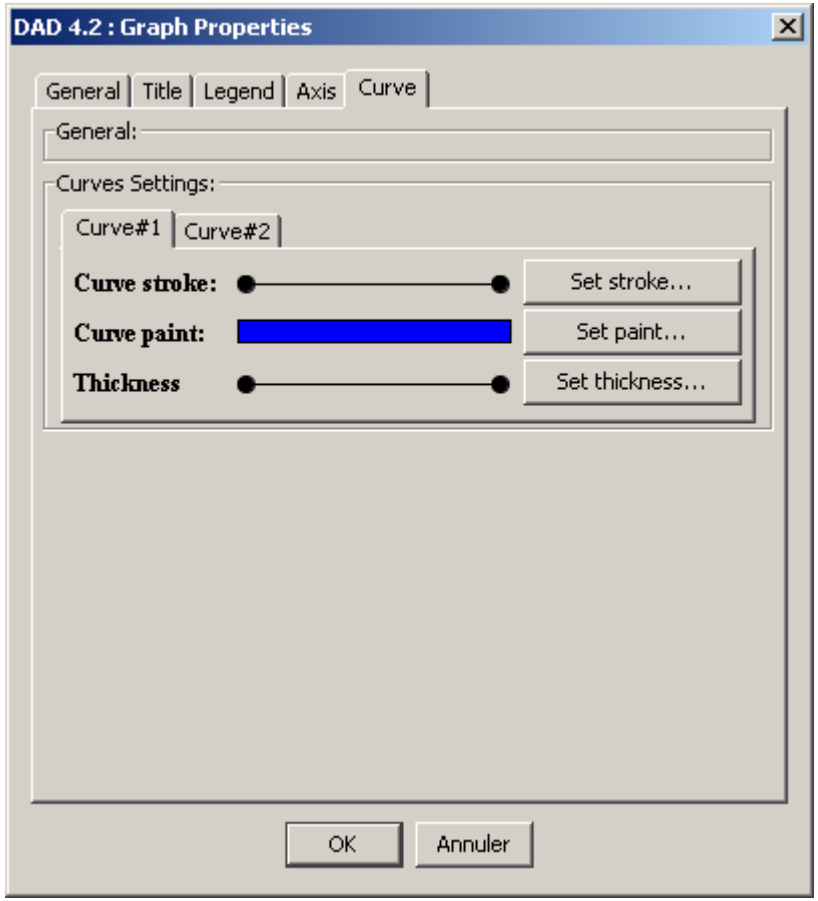


Select the desired stroke and click on the button OK to confirm your selection.

Curve Thickness: To choose the thickness of a giving curve, click on the button **Set Thickness**. The following widows appear:



Select the desired thickness, and click on the button OK to confirm your selection.



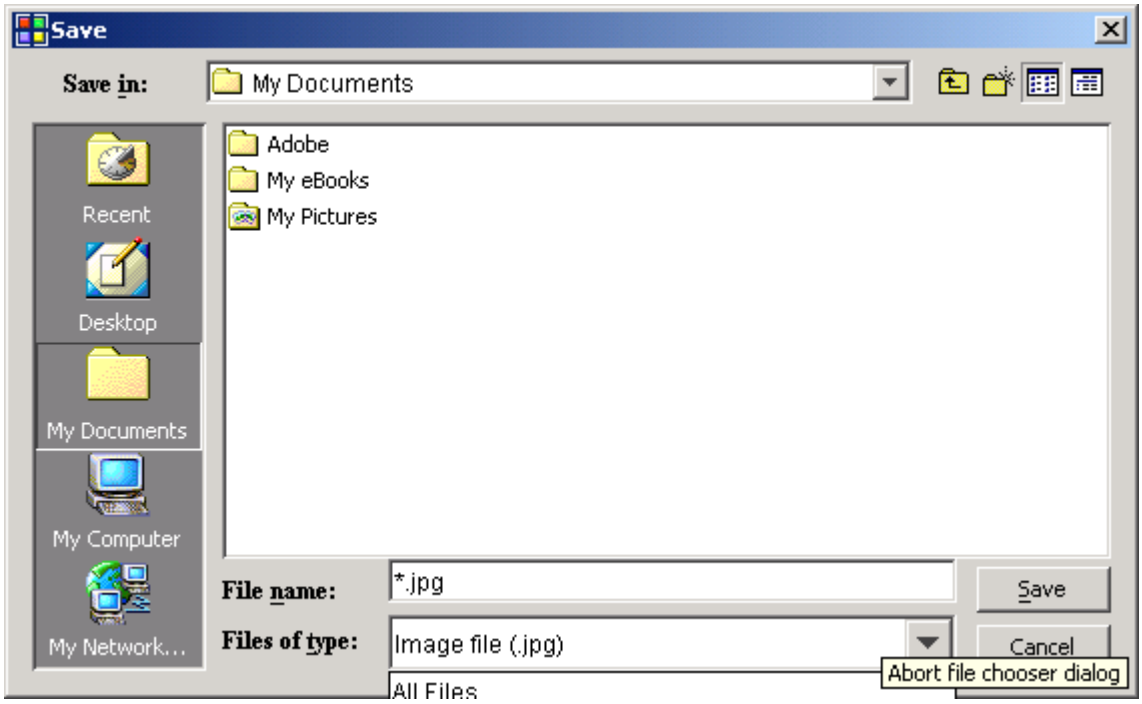
Curve Paint: To choose the colour of a giving curve, click on the button **Set Paint** and choose the new colour.

Saving graphs

You can save and use graphs in many popular text processors (including Word and Excell). The available formats are:

Extension	Description
*.png	Portable Network Graphic
*.jpg	JPEG File Interchange Format
*.pdf	Portable Document Format
*.ps	Postscript
*.tif	Tag Image File Format
*.bmp	Bitmat Image File

To save a graph made in DAD, select: File⇒Save and select the format by selecting the extension of the file.



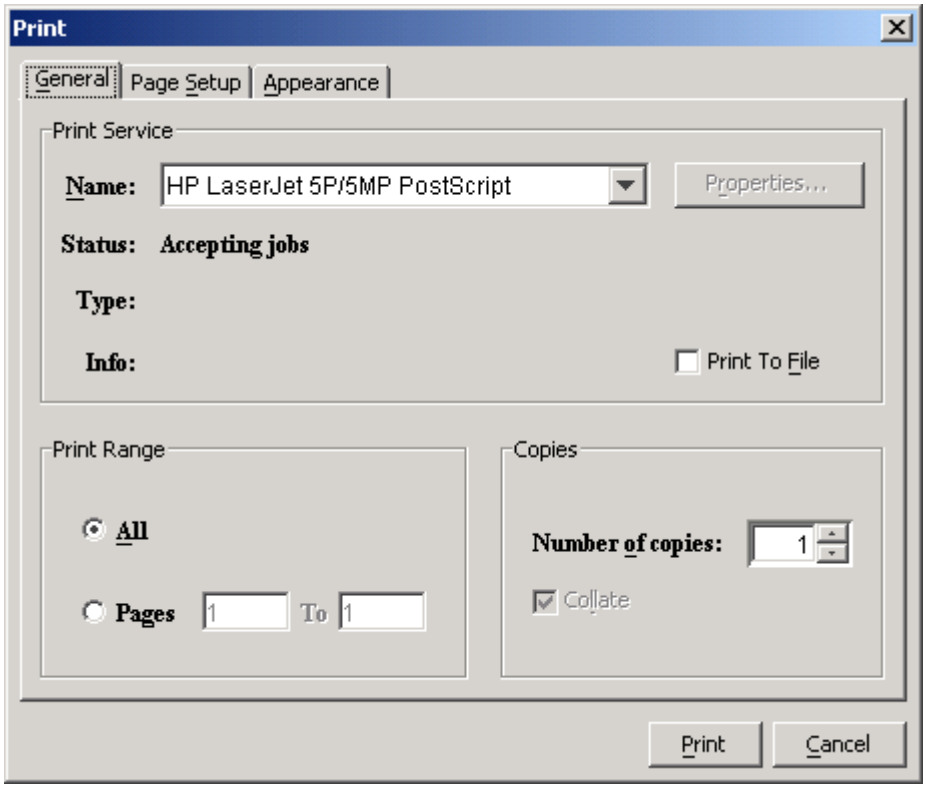
Saving coordinates of curves

To save the graph coordinates in ASCII format, select "File ⇒Save coordinates". The generated ASCII file takes the following format:

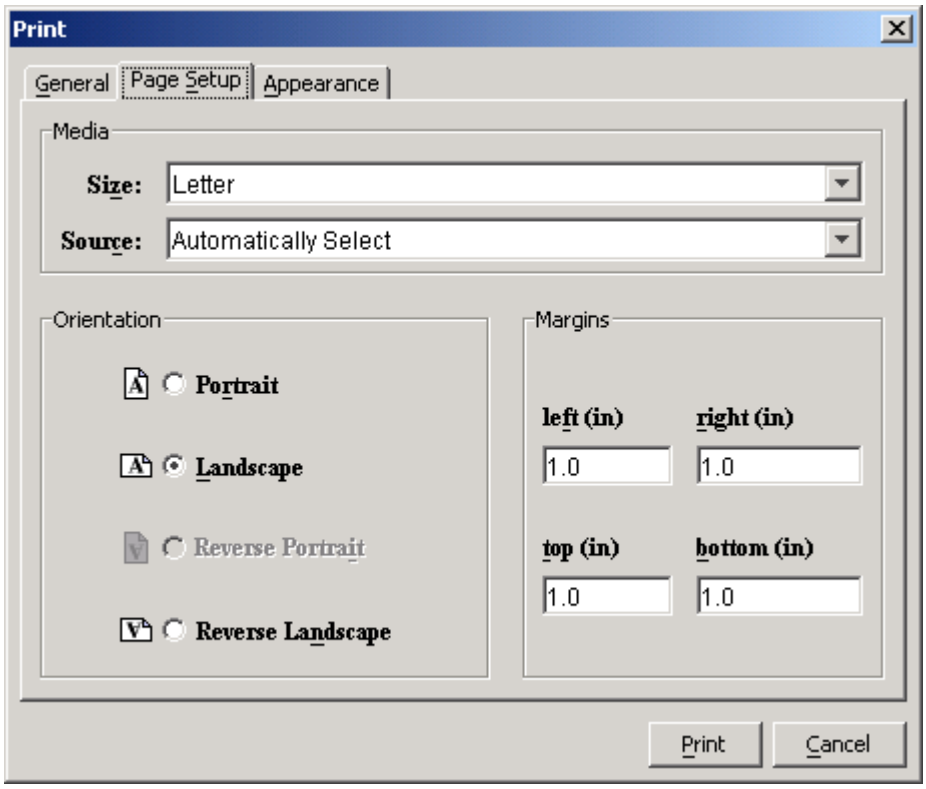
$$\overbrace{X1\quad Y1}^{\text{Curve1}}\ \overbrace{X2\quad Y2}^{\text{Curve2}}\cdots\text{etc}\cdots$$

Printing graphs

To print a graph, select “File ⇒Print”. The following windows appears:

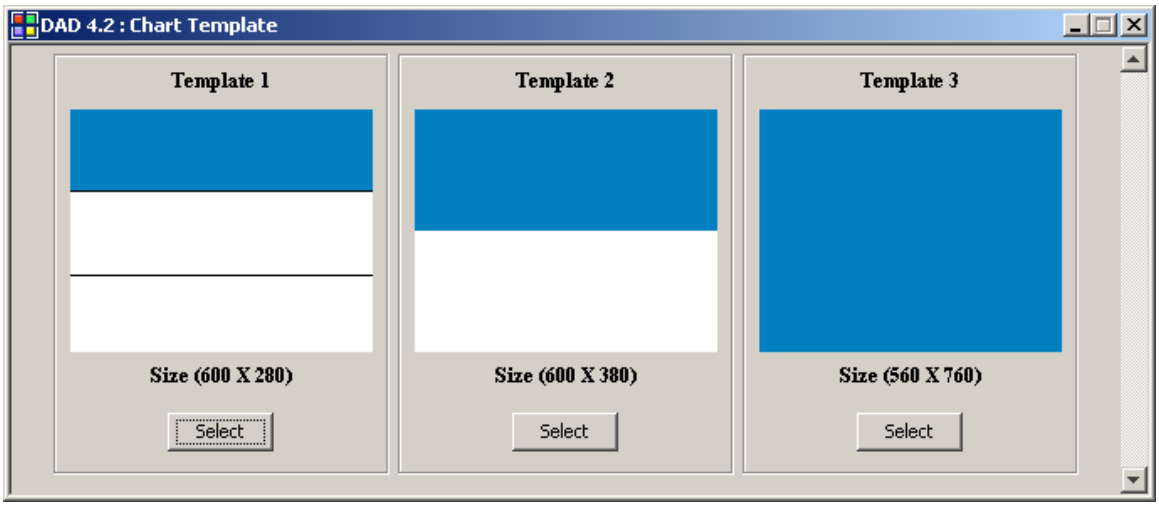


Select the desired Printer. To change orientation or margins, select “Page Setup”. When the following window appears, select the desired orientation and margins.



Templates

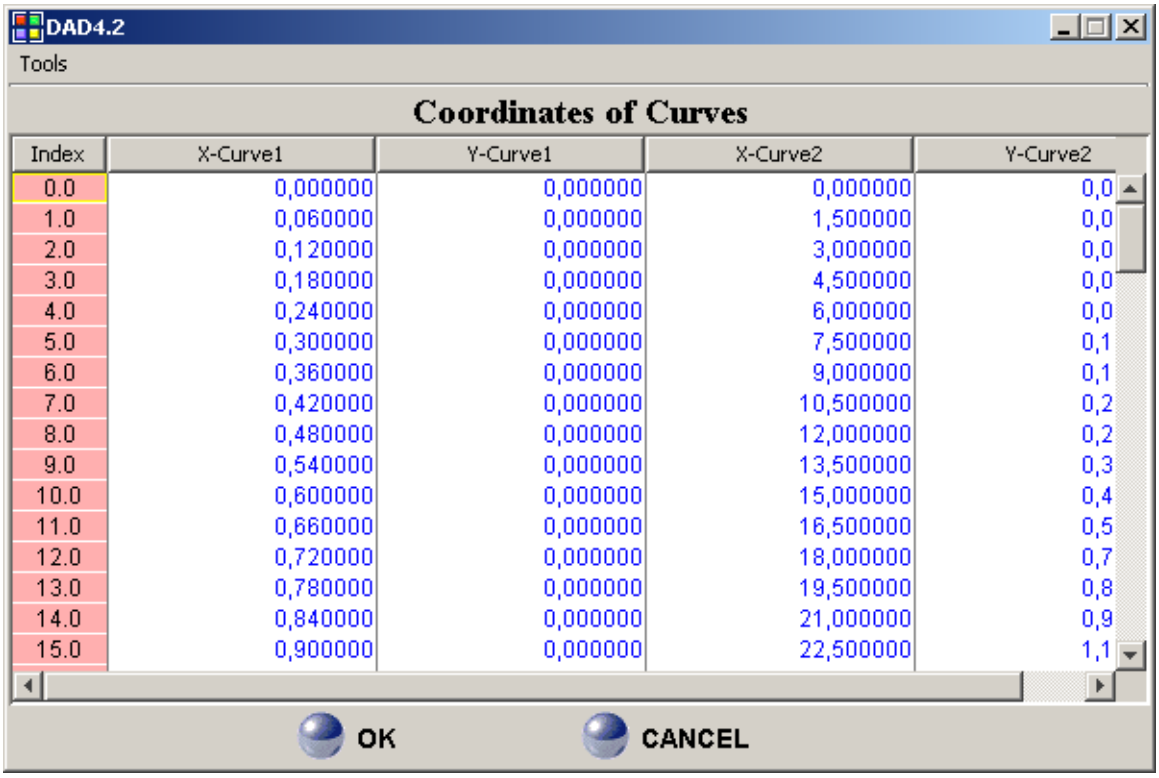
You can select one of DAD's several graphical templates to change the properties of a graph. These templates only use black and white colours. To select a template, select "Edit ⇒Templates". The following window appears:



- Template 1 can be inserted within a third of a page of a Word document.
- Template 2 can be inserted within half a page of a Word document.
- Template 3 can be inserted within a page of a Word document, with landscape orientation.

Editing coordinates

To edit coordinates of curves, select “Edit ⇒Edit Coordinates”. The following windows appears:



You can change the decimal number by using the item “Tools”. To close this window, click on the button “OK”.

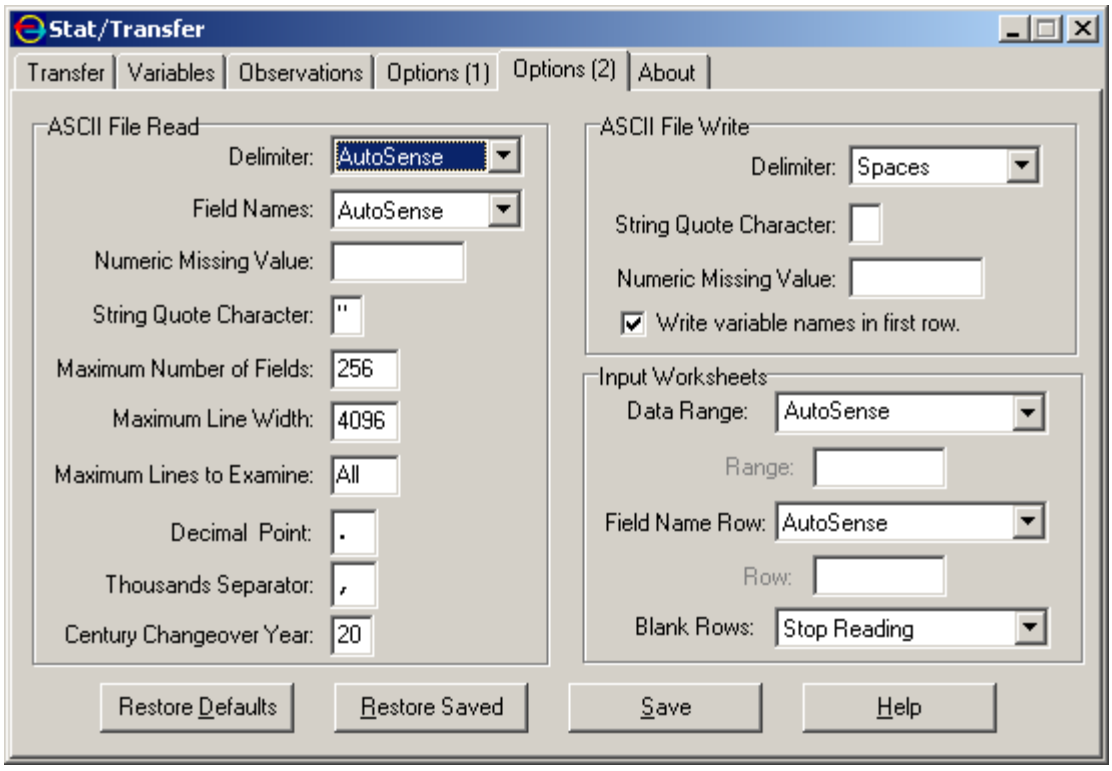
Preparing DAD ASCII Files in .daf Format with Stat/Transfer

A useful tool to produce DAD Ascii Format ("DAF") files is Stat/Transfer:
http://www.stattransfer.com/

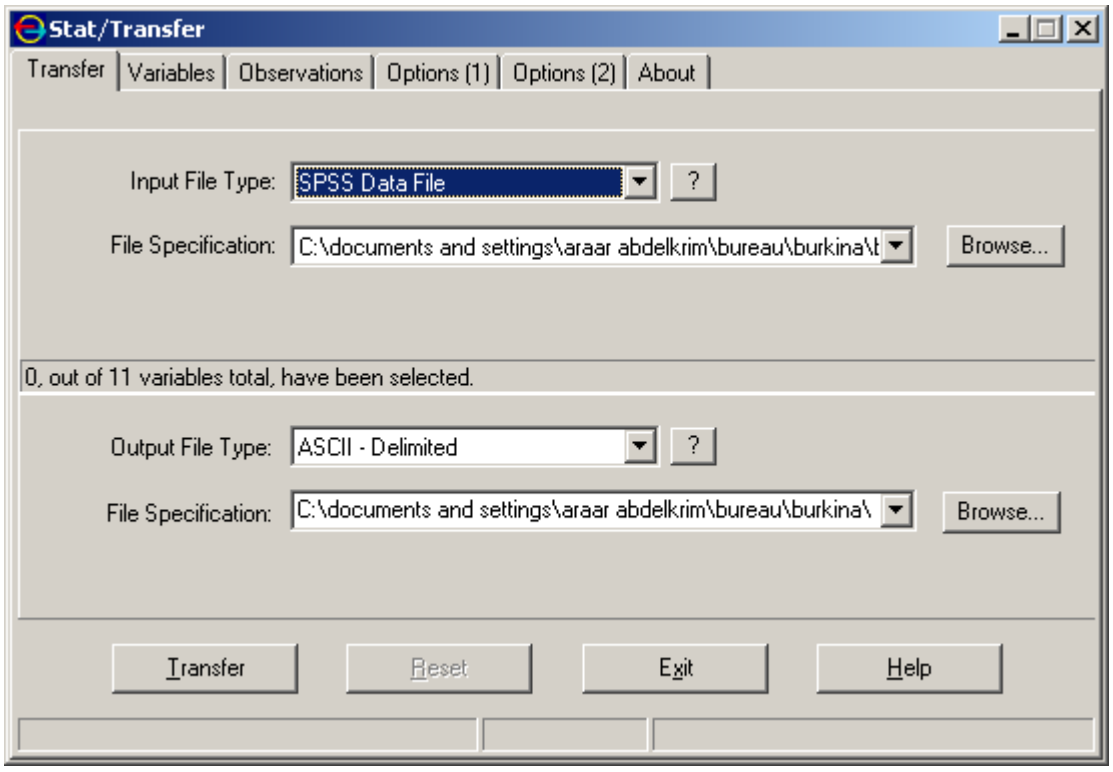
The following steps explain how one can prepare DAF files from any other format.

1. After opening Stat/Transfer, select from the main menu the item **"Option (2)".**
- **2.1** In the field ***ASCII File Writer***, select the Delimiter: Spaces.
- **2.2** Select the option ***Write variable names in first row***.

To do this only once, click on the button **"Save"** to save these preferences.



2. The usual next step is to select the item **"Transfer".**
- **2.1** First, select the type of the input file (SPSS. EXCEL…)
- **2.2** By using **"Browse"**, indicate the location of the input file.

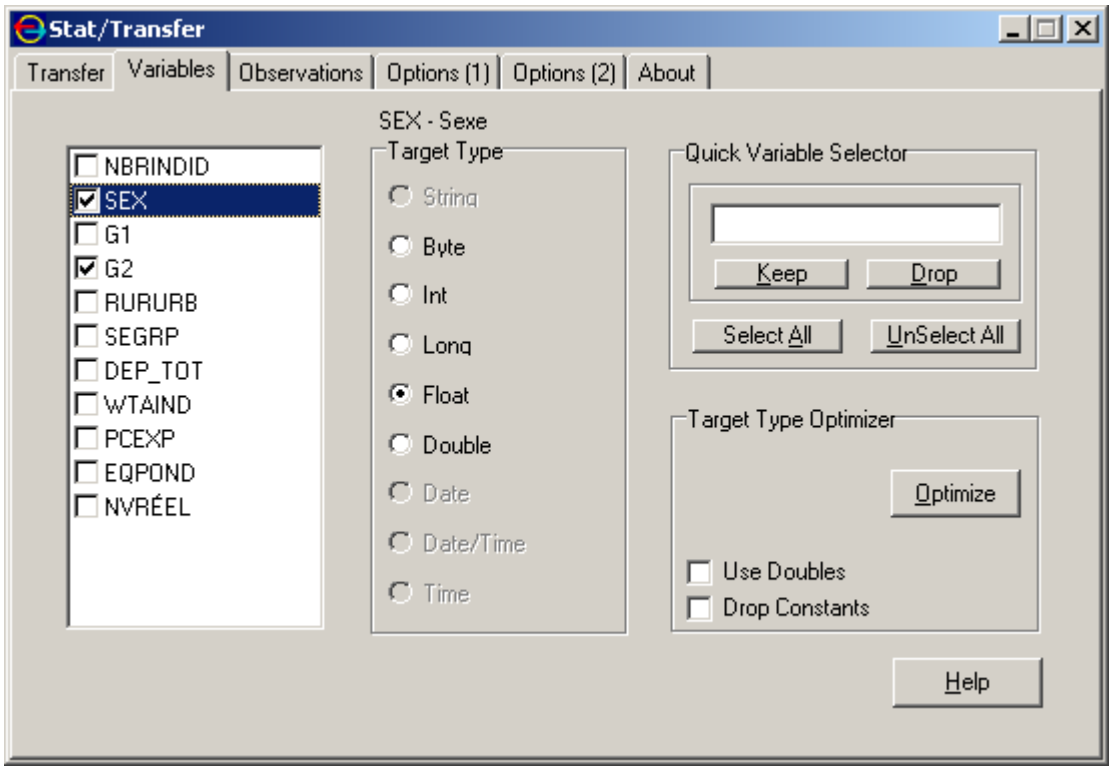


2.3 Select "***ASCII – Delimited***" as the type of output file.

2.4 By using **"Browse"**, indicate the location of the output file and write name with extension .daf. For example; the name is: Data1.daf

2.5 Click on the Button **"Transfer"** to produce the new file.

If you wish to save only some selected vectors in the DAF file, after step 2.2, select the item "**Variables"** and select those vectors you wish to save in the new DAF file. After this, continue to steps 2.3 to 2.5.


Stat/Transfer
Transfer
Variables
Observations
Options (1)
Options (2)
About
SEX - Sexe
NBRINDID
SEX
G1
G2
RURURB
SEGRP
DEP_TOT
WTAIND
PCEXP
EQPOND
NVRÉEL
Target Type
String
Byte
Int
Long
Float
Double
Date
Date/Time
Time
Quick Variable Selector
Keep
Drop
Select All
UnSelect All
Target Type Optimizer
Optimize
Use Doubles
Drop Constants
Help